BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Claims 1, 2, 4, 5, 50 and 51 are pending.

Rejection of Claims 1, 2, 4, 5, 50 and 51 Under 35 U.S.C. §112, first paragraph

Claims 1, 2, 4, 5, 50 and 51 are rejected under 35 U.S.C. §112, first paragraph, for enablement and written description. In the office action dated June 4, 2003, the Examiner asserted

> The specification teaches ... that "[t]he exogenous polypeptide was an 86-mer fusion protein having tandem ligand recognition sequences, a variegated segment of thirteen amino acids serving as a template for potential EK recognition sequences, a factor Xa cleavage site, segments linking the foregoing domains and linking the N-terminus of gene III protein" and ... that the tandem ligand recognition sites are "a linear binding sequence, [and] a constrained streptavidin binding loop. The instant claims are now limited to a "ligand recognition sequence", not tandem ones, with no limitation as to a Xa cleavage site or gene III. The potential EK recognition sequence is 7 amino acids not 13.
>
> It is maintained that absent convincing proof to the contrary, these additional limitations in the example have some effect upon which sites will be cleaved by enterokinase and are important in defining an enterokinase cleavage protein.

In the advisory action, the Examiner has maintained this rejection asserting that "although applicants argue that ... it is taught that several synthetic peptides with no ligand recognition sequence, with the recited EK cleavage site and three amino acids to the C-terminus may be cleaved with EK, it is pointed out that only one of these sequences, the 4th one, meet the requirements of claim 1."

Applicants respectfully traverse this rejection. Even if there is only one peptide demonstrated in the application that meets the requirements of claim 1, has none of the structures required by the Examiner and is shown to be cleaved by enterokinase, this is still evidence that in the absence of a tandem ligand recognition sequences, Xa cleavage sites or gene III, the claimed fusion peptides will be cleaved by enterokinase. In addition, the Examiner completely disregards the additional evidence provided by Applicants in the last response that demonstrates a fusion

protein having an enterokinase cleavage site but none of the other structures required by the Examiner and that was repeatably cleaved in the presence of low levels of enterokinase. In other words, this was an enterokinase cleavage site flanked by entirely different structures than required by the Examiner and it was cleaved by enterokinase. The example relied upon by the Examiner together with the two examples discussed herein show that three very different constructs can be cleaved. Thus, Applicants have provided the Examiner with evidence that the claimed fusion polypeptides having the recited structure will be cleaved without the need for a tandem ligand recognition sequence, Xa cleavage sites or gene III. The Examiner, on the other hand, has provided absolutely no evidence that suggest that these additional structures are needed for cleavage by enterokinase.

To further support the fact that a skilled artisan would view the evidence provided as sufficient to describe the claimed invention and enable a skilled artisan to make and use the claimed polypeptides, Applicants submit herewith a Declaration Under 37 CFR §1.131 by Dr. Robert Ladner (hereafter referred to as "the Declaration"). As provided in the Declaration, evidence has been provided that the claimed enterokinase cleavage motif is cleaved regardless of what is coupled to either end. A skilled artisan looking at this evidence would believe that other peptides having the claimed enterokinase cleavage motif would behave in the same manner. As discussed in the Declaration, much is known about the naturally occurring enterokinase cleavage site and it has been successfully cleaved in countless applications regardless of the sequences on either side of it. The structure of the claimed enterokinase cleavage site is of a similar length and structure as the naturally occurring enterokinase cleavage site. In addition, the naturally occurring enterokinase site has aspartic acid-lysine amino acids that are known to play an important role in cleavage of that site by enterokinase by formation of a hydrogen bond between the lysine residue and the enzyme pocket. In the present application, the enterokinase cleavage sites have aspartic acid-arginine amino acids. It has been shown that the claimed enterokinase cleavage sites are more efficiently cleaved than the naturally occurring enterokinase cleavage site. This is perhaps due to the structure of arginine which allows for more hydrogen bonds to be formed with the enzyme pocket. Thus, the naturally occurring enterokinase cleavage site and the

claimed enterokinase cleavage sites appear to work by the same mechanism. Thus, based upon the knowledge in the art regarding enterokinase cleavage sites, the close resemblance of all of the claimed enterokinase cleavage sites with the naturally occurring enterokinase cleavage site and the evidence provided in the application, a skilled artisan would recognized that Applicants were in possession of the claimed fusion polypeptides and would have reason to believe that the claimed fusion polypeptides would result in enterokinase cleavage without the additional structures recited by the Examiner. It is clear that Applicants have provided sufficient evidence to support the fusion polypeptides as currently claimed. The burden is on the Examiner to provide any evidence to support his assertions. In the absence of such evidence, Applicants respectfully request that the Examiner withdraw this rejection.

The Examiner indicated that Applicants response regarding the justification for the general formula provided by the claims has merit and therefore, is not addressed in this reply.

Enclosed is a check for the Petition for Extension of Time fee. Please apply any other charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: 4/29/04

Laurie Lawrence

Laurie Butler Lawrence
Reg. No. 46,593

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110-2804
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ley et al. Art Unit : 1652
Serial No. : 09/884,767 Examiner : C. Patterson
Filed : June 19, 2001
Title : NOVEL ENTEROKINASE CLEAVAGE SEQUENCES
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF ROBERT LADNER, Ph.D. UNDER 37 CFR §1.132

I, Robert Ladner, Ph.D., pursuant to 37 C.F.R. § 1.132, declare the following:

1. My educational and professional experience and qualifications are presented in the attached Curriculum Vitae.

2. I am currently employed by Dyax Corp. as the Senior Vice President and Chief Scientific Officer.

3. I have reviewed the Office Action mailed June 4, 2003 and the Advisory Action mailed December 11, 2003 regarding the above-referenced application and understand that the Examiner has rejected the claims asserting that

> the instant claims are now limited to a 'ligand recognition sequence', not tandem ones, with no limitation as to a Xa cleavage site or gene III. The potential EK recognition sequence is 7 amino acids not 13. It is maintained that absent convincing proof to the contrary, these additional limitations in the example have some effect upon which sites will be cleaved by enterokinase and are important in defining an enterokinase cleavage protein.

4. Contrary to the Examiner's assertions, the present application demonstrates that a peptide having the recited structure will be cleaved without the presence of these additional elements. Specifically, at page 40 of the present application, a synthetic peptide was made that had no ligand recognition sequence (Z_1), had the recited EK cleavage site of Xaa_1-Xaa_2-Xaa_3-Xaa_4-Asp-Arg, and had three amino acids making up the Xaa_5-Z_2 portion of the fusion protein. It was demonstrated that this synthetic peptide was completely cleaved by enterokinase between the Arg and Xaa_5. Thus, it is clear that without the additional elements recited by the Examiner, a fusion peptide having a

sequence covered by the claimed motif will be cleaved by enterokinase in the absence of a tandem recognition sequence, a Xa cleavage site and gene III.

5. In subsequent experiments, we have prepared a fusion protein having an enterokinase cleavage site but none of the other structures required by the Examiner and have show that this fusion protein was repeatably cleaved in the presence of low levels of enterokinase. The fusion protein used in these experiments included a prosequence, a Kex2 cleavage site, a 6-mer enterokinase cleavage site (DINDDR), a serine residue, and a kunitz domain. We demonstrated that by using the recited enterokinase cleavage site a kunitz protein could be repeatable cleaved from the fusion protein without any impurities. Thus, it is clear that the enterokinase cleavage site as recited in the claims is sufficient to result in the cleavage of a fusion protein without the presence of a tandem ligand recognition sequence, a Xa cleavage site or a gene III sequence. In addition, it is clear that the 6-mer structure recited in the claims is sufficient for enterokinase cleavage.

6. A person of ordinary skill in the art would expect fusion proteins having other peptide sequences covered by the claimed motif to be cleaved by enterokinase in the absence a tandem ligand recognition sequence, a Xa cleavage site or a gene III sequence. It has been well documented that the natural enterokinase cleavage site having the sequence Asp-Asp-Asp-Asp-Lys is cleaved by enterokinase regardless of what is upstream or downstream from the cleavage site. In fact, the naturally occurring enterokinase cleavage site has been used as part of various fusion proteins and is cleaved by enterokinase in each of these different settings. See, for example, Yamaguchi et al. (2002) *J. Biol. Chem.* 277(9):6806-6812 (which discloses the production of active spinesin by expression of mature spinesin preceded by the enterokinase cleavage site and subsequent cleavage with enterokinase); Hung et al. (1992) *J. Biol. Chem.* 267: *J. Biol. Chem.* 267(29):20831-20834 (which discloses the production of a thrombin receptor by expression of the enterokinase cleavage site and the thrombin receptor, followed by subsequent treatment with enterokinase); Gasmi et al. (1999) *Biochem. J.* 344(Pt. 2):331-337 (which discloses the production of YSA1H by expression of YSA1H preceded by the enterokinase cleavage site and subsequent cleavage with enterokinase); Karlson et al. (2002) J. Biol. Chem. 277(21):18579-18585 (which discloses the production of MCP-4

by expression of MCP-4 preceded by the enterokinase cleavage site and subsequent cleavage with enterokinase), submitted herewith as Exhibits B-E, respectively.

7. In view of the evidence provided in the application and in this declaration, there is no reason to believe that peptides having the claimed peptide cleavage site motif would function differently than the natural enterokinase cleavage site by requiring specific upstream or downstream sequences in order to be cleaved by enterokinase. The claimed peptide motif is of similar length and structure to that of the naturally occurring enterokinase cleavage site. In addition, it is known in the art that the lysine residue in the naturally occurring enterokinase cleavage site forms a hydrogen bond with the S1 pocket of enterokinase, thereby aiding in the cleavage of this peptide by enterokinase. As demonstrated in the application, a peptide having the claimed motif demonstrates more efficient hydrolysis than the naturally occurring enterokinase cleavage site. This may be due in part to the presence of amino acids Asp-Arg in the claimed peptides versus the Asp-Lys in the naturally occurring enterokinase cleavage site. It is well known in the art that arginine provides a larger side group that can create more hydrogen bonds than lysine. Thus, the presence of the arginine may form more bonds with the S1 pocket of the enterokinase thereby aiding in more efficient cleavage by the enzyme. In view of the knowledge in the art of how the naturally occurring enterokinase site interacts with enterokinase, the requirement that the claimed peptides have the Asp-Arg sequence, and the evidence that a peptide having the recited Asp-Arg was more efficiently cleaved by enterokinase, it is reasonable to believe that any peptide having the claimed motif would be cleaved by enterokinase regardless of the sequences occurring upstream or downstream from the cleavage site. The evidence provided in the application and in this declaration support this conclusion.

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

DATE: 27 April 2004

Robert Ladner

Robert Ladner, Ph.D.

20847106.doc

EXHIBIT B

ISSN 0021-9258 (print)
ISSN 1083-351x (electronic)
JBCHA3 277(9) 6759-7628 (2002)

The Journal of Biological Chemistry

MARCH 1, 2002 VOLUME 277 NUMBER 9

PUBLISHED BY THE AMERICAN SOCIETY FOR
BIOCHEMISTRY AND MOLECULAR BIOLOGY

Founded by Christian A Herter and Sustained in Part by the Christian A Herter Memorial Fund

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 277, No. 9, Issue of March 1, pp. 6806–6812, 2002
Printed in U.S.A.

Spinesin/TMPRSS5, a Novel Transmembrane Serine Protease, Cloned from Human Spinal Cord*

Received for publication, April 24, 2001, and in revised form, December 5, 2001
Published, JBC Papers in Press, December 12, 2001, DOI 10.1074/jbc.M103645200

Nozomi Yamaguchi‡§, Akira Okui¶, Tatsuo Yamada‖, Hiroshi Nakazato, and Shinichi Mitsui‡**

*From the ‡Department of Cell Biology, Research Institute for Neurological Diseases and Geriatrics, Kyoto Prefectural University of Medicine, Kyoto 602-8566, Japan, the ¶Research and Development Center, Fuso Pharmaceutical Co., Morinomiya, Joto-ku, Osaka 536-8523, Japan, the ‖Department of Internal Medicine and Health Care, Fukuoka University, Hakata 814-0180, Japan, and the **Center for Applied Toxicology, Avenida Vital Brazil 1500, 05503-900, Sao Paulo, SP, Brazil*

A cDNA encoding a novel serine protease, which we designated spinesin, has been cloned from human spinal cord. The longest open reading frame was 457 amino acids. A homology search revealed that the human spinesin gene was located at chromosome 11q23 and contained 13 exons, the gene structure being similar to that of TMPRSS3 whose gene is also located on 11q23. Spinesin has a simple type II transmembrane structure, consisting of, from the N terminus, a short cytoplasmic domain, a transmembrane domain, a stem region containing a scavenger receptor-like domain, and a serine protease domain. Unlike TMPRSS3, it carries no low density lipoprotein receptor domain in the stem region. The extracellular region carries five *N*-glycosylation sites. The sequence of the protease domain carried the essential triad His, Asp, and Ser and showed some similarity to that of TMPRSS2, hepsin, HAT, MT-SP1, TMPRSS3, and corin, sharing 45.5, 41.9, 41.3, 40.3, 39.1, and 38.5% identity, respectively. The putative mature protease domain preceded by H_6DDDDK was produced in *Escherichia coli*, purified, and successfully activated by immobilized enterokinase. Its optimal pH was about 10. It cleaved synthetic substrates for trypsin, which is inhibited by *p*-amidinophenylmethanesulfonyl fluoride hydrochloride but not by antipain or leupeptin. Northern blot analysis against mRNA from human tissues including liver, lung, placenta, and heart demonstrated a specific expression of spinesin mRNA in the brain. Immunohistochemically, spinesin was predominantly expressed in neurons, in their axons, and at the synapses of motoneurons in the spinal cord. In addition, some oligodendrocytes were clearly stained. These results indicate that spinesin is transported to the synapses through the axons after its synthesis in the cytoplasm and may play important roles at the synapses. Further analyses are required to clarify its roles at the synapses and in oligodendrocytes.

Serine proteases have essential functions in biological processes such as the activation of complement and blood coagulation. Recently some serine proteases have been reported to contain a transmembrane domain that anchors the protease molecule to the cell membrane. During the last few years, many type II transmembrane serine proteases (TTSPs, referred to in this article as TMPRSS)[1] from mammals have been cloned and reported, namely enterokinase (1), hepsin (2), HAT (3), corin (4, 5), MT-SP1 (epithin) (6), matriptase (7), TMPRSS2 (epitheliasin) (8, 9), TMPRSS3 (10, 11), seprase (12), TADG12 (13), and TADG15 (14).

The common structural features of TMPRSSs are that they contain, from the N terminus, a short cytoplasmic domain, a transmembrane domain, a stem region, and a serine protease domain, the latter two being outside of the cell. The stem region varies in length and contains various modulatory domains. The length of these proteases ranges from 400 to over 1000 amino acid residues. The longest is corin at 1042 residues (4), and the shortest is hepsin at 417 residues. Hepsin has the simplest domain structure, having no unique modulatory domain in the stem region (2). On the other hand, enterokinase at a length of 1019 residues has the most complicated multiple domains in the stem region, *i.e.* a SEA (sea urchin sperm protein-enterokinase-agrin) domain; two low density lipoprotein receptor class A domains; two CUB (Cls/Clr, urchin embryonic growth factor, and bone morphogenic protein 1) domains; a MAP (meprin, A5 antigen, and receptor protein phosphatase μ) domain; and an SRCR (scavenger receptor cysteine-rich) domain (1).

At present, the roles of these domains have not been clarified, although the presence of a cytoplasmic domain suggests involvement in intracellular signaling. The various domains in the stem region may function in the recognition of other molecules, *e.g.* proteolytic substrates and inhibitors as well as other proteins and ligands, soluble or matrix-bound, on other cells, suggesting important roles for TMPRSSs in the body (for a review, see Ref. 15). Corin and matriptase process atrial natriuretic peptide (16) and hepatocyte growth factor (17), respectively. Enterokinase has long been known to have an essential role in the processing of digestive proteases (18), while trypsin is

We have been studying the brain-specific serine proteases and have newly cloned and characterized neurosin/PRSS9 (19), hippostasin/PRSS20 (20, 21), and motopsin/PRSS12 (22, 23). Neurosin and hippostasin, whose genes are found on chromosome 19q13.3, are secreted and belong to the kallikrein-like serine protease family. Motopsin, whose gene is located on

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the DDBJ/GenBank™/EBI Data Bank with accession number(s) AB028140.

§ To whom correspondence should be addressed. Tel.: 81-75-251-5797; Fax: 81-75-251-5848; E-mail: nozomi@koto.kpu-m.ac.jp.

[1] The abbreviations used are: TMPRSS, transmembrane protease serine (or TTSP, type II transmembrane serine protease); CNS, central nervous system; CSF, cerebrospinal fluid; HAT, human airway trypsin-like protease; KLH, keyhole limpet hemocyanin; MT-SP1, membrane-type serine protease 1 (also known as matriptase); RACE, rapid amplification of cDNA ends; MCA, 4-methyl-coumaryl-7-amide; Boc, *N*-*tert*-butoxy-carbonyl; Bz, benzoyl; Z, benzyloxycarbonyl.

chromosome 5, has a unique and complicated structure similar to TMPRSS, including, from the N terminus, a proline-rich domain, a kringle domain, three scavenger receptor cysteine-rich domains, and a protease domain. However, motopsin has a putative signal sequence at the N terminus without an obvious hydrophobic transmembrane domain and thus would appear to be a secreted protease.

As part of our continuing efforts to characterize serine proteases from the CNS, we have cloned from a human spinal cord mRNA pool a TMPRSS that we designated spinesin or TMPRSS5. As far as we know, this is the first report of a TMPRSS identified in the CNS.

EXPERIMENTAL PROCEDURES

Materials—Human tissues of the CNS for immunohistological analyses were obtained with informed consent within 12 h of death. Cerebrospinal fluids (CSFs) were obtained with informed consent from patients with non-central nervous system diseases. Human brain mRNA, multiple tissue Northern blots, and tissue extracts were purchased from CLONTECH (Palo Alto, CA). Tissue culture media, supplements, pTrcHisB vector, and competent *Escherichia coli* DH5α cells were from Invitrogen. All other chemicals were obtained from Wako Chemicals (Osaka, Japan).

Antibodies—Rabbit polyclonal antibodies for Western blotting and immunohistochemical analyses were raised against two KLH-conjugated peptides, KLH-CSEASAEEALLP (anti-human spinesin A) and KLH-CAGLVSHSAVRPHQG (anti-human spinesin B), and purified using protein A-Sepharose (Amersham Biosciences, Inc.). The former peptide sequence is derived from the stem region, and the latter was derived from the protease domain (see Fig. 1).

Isolation of Human Spinesin cDNA Clones—Poly (A)$^+$ RNA from human CNS (CLONTECH) was reverse-transcribed by using the SuperScript Preamplification System (Invitrogen) according to the instruction manual. PCR with a pair of degenerate primers, DP-S and DP-A (Table I), was performed as described previously (22, 24). The PCR products were ligated into the pGEM-T Easy vector (Promega, Madison, WI), cloned, and sequenced using an automatic sequencer (DSQ-1000, Shimadzu Co., Kyoto Japan). A clone carrying a 465-bp fragment was found to have a novel serine protease-related sequence. Based on this sequence, specific primers were synthesized for the rapid amplification of cDNA ends (RACE, Table I). For 3′-RACE, human CNS poly(A)$^+$ RNA was reverse-transcribed using oligo(dT) with an adaptor primer sequence at the 5′-end, $TGGAAGAATACGCGGCCGCAGT_{17}$. The cDNA was then amplified using forward primer 1 and the adaptor primer, the products of which were further amplified by nested PCR using primer 2 and the adaptor primer. 5′-RACE was performed using a Marathon cDNA amplification kit (CLONTECH) according to the instruction manual. In brief, nested PCR with AP2 and primer 3 was performed using products of PCR with primer 4 and AP1 as a template.

Northern Blot Hybridization—Northern blot hybridization against human multiple tissues was carried out using a commercially available membrane (CLONTECH). The cDNA carrying the full-length human spinesin open reading frame amplified using primers 5 and 6 was labeled by the random labeling method using a Takara BcaBEST labeling kit (Takara Shuzo Co. Ltd., Shiga, Japan). Hybridization was carried out in ExpressHyb hybridization solution (CLONTECH) at 60 °C overnight, and the final wash was performed in 0.1× saline/sodium phosphate/EDTA containing 0.1% SDS at room temperature for 10 min. The radioactivity was detected using an FLA-2000 image analyzer (Fuji Photo Film Co. Ltd., Tokyo, Japan).

Immunohistochemical Analysis—The CNS including spinal cord from a non-neurological patient (65-year-old Japanese male) was obtained 2–12 h after death. Small blocks were dissected and fixed in 0.1 M phosphate-buffered 4% paraformaldehyde for 2 days and then stored in 0.1 M phosphate-buffered saline containing 15% sucrose and 0.1% sodium azide and kept at −70 °C until use. Sections were cut on a cryostat at 20-μm thickness and washed in phosphate-buffered saline. The antibody, diluted 1:2000 with phosphate-buffered saline-Tween, was incubated with the specimens at 4 °C for 48 h. After a wash with phosphate-buffered saline-Tween, the slides were incubated with alkaline phosphatase-labeled goat anti-rabbit IgG for 60 min at room temperature. After another wash, immunoreactivity was visualized with 4-nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl phosphate. Counter staining was not performed.

Expression and Purification of Recombinant Spinesin in a Bacteria Expression System—To obtain an active recombinant spinesin, a cDNA fragment encoding the putative mature enzyme of spinesin (Ile^{218}–Leu^{457}) was amplified by PCR (forward primer, ATAGTTGGTGGGCAGTCTGT; reverse primer, primer 6 in Fig. 1) and subcloned into pTrcHisB between the *Bam*HI site, which had been treated with mung bean exonuclease following the instructions of the manufacturer, and the *Eco*RI site. The resultant vector carrying chimera cDNA encoding H_6DDDDK-(^{218}I–L^{457}) was transformed into DH5α, and the recombinant protein was induced using 0.7 mM isopropyl-β-D-thiogalactopyranoside. The recombinant protein in the cells from 100 ml of culture, mostly in inclusion bodies, was collected and suspended in 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl and 1% Triton X-100. The suspension (5 ml) was sonicated, and inclusion bodies were collected by centrifugation and resuspended. After three rounds of sonication and centrifugation, the final pellet was dissolved with 5 ml of 8 M urea in the same buffer without Triton X-100 by sonication and shaken at room temperature for 1 h. Then the solution was diluted 10 times with 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl under vigorous stirring and centrifuged for 30 min at 3,500 rpm to remove debris. The supernatant was applied to a Talon column (1 × 1 cm) equilibrated with 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl. Following a sufficient wash (20 ml) with the same buffer containing 15 mM imidazole, spinesin was eluted by 100 mM imidazole in 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl. After the removal of imidazole using a PD-10 desalting column (1 × 5 cm) equilibrated with 20 mM Tris-HCl buffer, pH 8.0, containing 0.2 M NaCl, spinesin was activated by incubation with recombinant enterokinase (EK Max, Invitrogen) immobilized on *N*-hydroxysuccinimide-Sepharose (1 ml) (Amersham Biosciences, Inc.) for 30 min at room temperature to remove the H_6DDDDK sequence.

Enzyme Assay—Five μl of the enzyme activated by the enterokinase column was incubated with 100 μl (20 μM) of various synthetic peptide substrates, *i.e.* Boc-Gln-Ala-Arg-MCA, Boc-Phe-Ser-Arg-MCA, Bz-Arg-MCA, Boc-Val-Leu-Lys-MCA, Pyr-Gly-Arg-MCA, Pro-Phe-Arg-MCA, Boc-Val-Pro-Arg-MCA, Z-Arg-Arg-MCA, Arg-MCA, or Z-Phe-Arg-MCA (Peptide Inst. Inc., Osaka, Japan) in 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl at 37 °C. After a 30-min incubation, the fluorescence (excitation at 380 nm, emission at 460 nm) was measured using a plate reader (Cytofluor 2300, Millipore, Bedford, MA). The effect of pH on the activity of spinesin treated with 0.01 unit of recombinant enterokinase was tested using Boc-Gln-Ala-Arg-MCA as a substrate in either 0.1 M phosphate buffer or 0.1 M Tris-HCl buffer that contained 0.2 M NaCl. The reaction was carried out under the conditions described above. An inhibitor profile was obtained by preincubating for 30 min at 37 °C with a final concentration of 1 μM *p*-amidinophenylmethanesulfonyl fluoride hydrochloride, 1 mM leupeptin, or 1 mM antipain. The remaining enzyme activity was expressed relative to a control value obtained by adding buffer without inhibitor.

Zymography—Gelatin or casein (270 μg/ml) was copolymerized in a standard 12.5% SDS-polyacrylamide gel. The activated recombinant spinesin (100 ng) was electrophoresed at a constant current of 20 mA under nonreducing condition. The gel was washed with 20 mM Tris-HCl (pH 8.0), 0.2 M NaCl containing 1% Triton X-100 at 37 °C for 3 h and then incubated in 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl at 37 °C overnight. The gel was stained with Coomassie Brilliant Blue.

Western Blot Analysis—The samples were applied to a 12.5% polyacrylamide gel containing SDS and electrophoresed. The separated proteins were transferred onto polyvinylidene difluoride membrane and then incubated overnight at room temperature with anti-human spinesin A or B diluted 2000-fold with 20 mM Tris-HCl, pH 7.4, containing 0.05% Tween 20 and 0.2 M NaCl. After a wash with the same buffer, the membrane was incubated with alkaline phosphatase-labeled goat anti-rabbit IgG for 60 min at room temperature. After another wash, immu-

TABLE I
PCR primers

Name	Sequence
DP-S	GTGCTCACNGCNGCBCAYTG
DP-A	AGCGGNCCNCCDSWRTCVCC
Primer 1	ACTGCTGCACATTGTATG
Primer 2	GCTCTCAACTTCTCAGACAC
Primer 3	AGGGGGCCCCCGCTATCTCC
Primer 4	ACTCAGCTACCTTGGCGTAGA
Primer 5	GCTTTACAACAGTGCTACTGAC
Primer 6	AAGGAATTCGAGGAAACAGCAGGACTCAGA
Adaptor primer	TGGAAGAATACGCGGCCGCAG
AP1	CCATCCTAATACGACTCACTATAGGGC
AP2	ACTCACTATAGGGCTCGAGCGGC

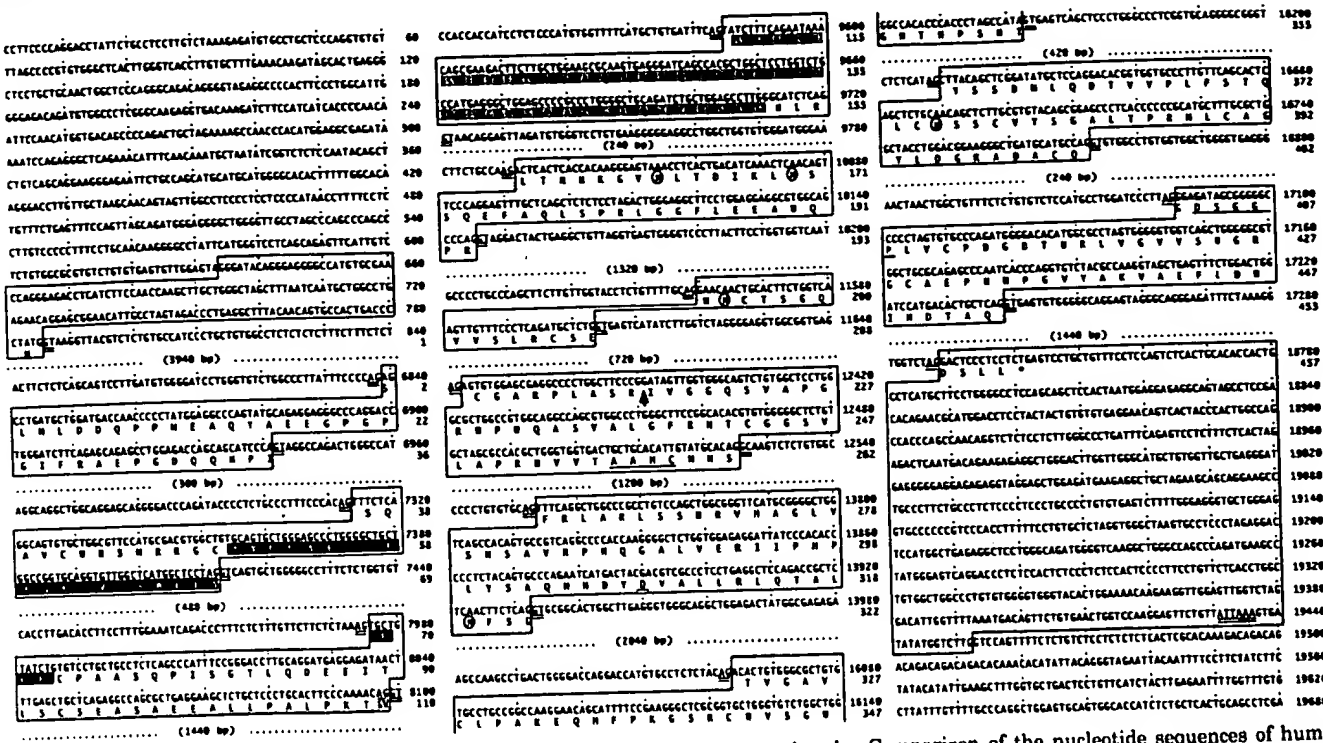


FIG. 1. **Structures of cDNA, deduced amino acids, and gene of human spinesin.** Comparison of the nucleotide sequences of human spinesin cDNA and human chromosome 11q clone DNA (GenBank™ accession no. AP002436) revealed the structure of the human spinesin gene. The 13 exons are *boxed*, and the intervening sequences are not shown except for the exon/intron boundaries. The exon/intron boundary consensus (GT or GC/AG) sequence and poly(A) signal are *double underlined*. Amino acids are *numbered* starting from the putative first initiating Met. A transmembrane and a scavenger receptor-like domain are indicated by *white* and *shaded letters*, respectively. The essential triad and the putative processing site are indicated by an *underline* and by an *arrow*, respectively. Putative *N*-glycosylation sites are *circled*.

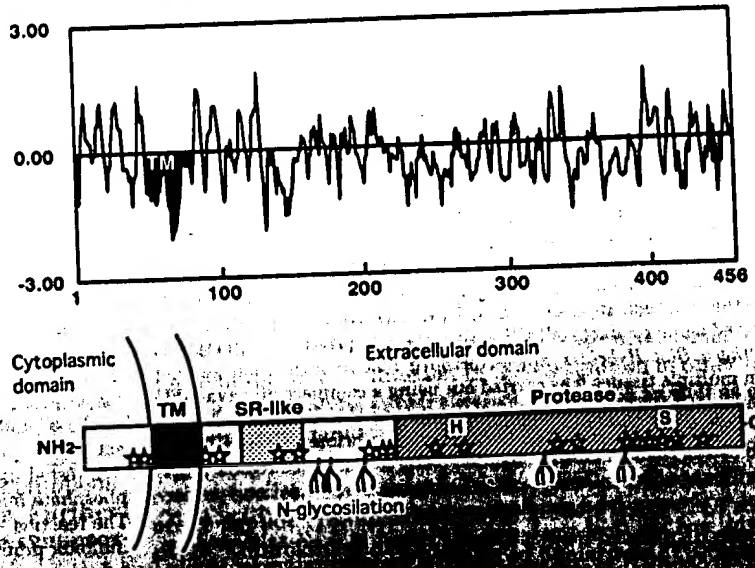


FIG. 2. **Hydropathy plot of the deduced amino acid sequence of spinesin.** The method of Hopp and Woods (25) was used with averaging over a window of 10 residues. Hydrophobic residues show negative values, whereas hydrophilic residues show positive values. The structure of spinesin is illustrated *under* the plot. *TM*, transmembrane domain; *SR-like*, scavenger receptor-like domain; ☆, cysteine residues; *upside-down* ψ, *N*-glycosylation sites.

noreactivity was visualized with 4-nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl phosphate.

RESULTS

Nucleic Acid Sequence of Human Spinesin cDNA—PCR using degenerate primers designed from serine protease motifs, AAHC and DSGGP, amplified a 465-bp fragment from a cDNA library of human spinal cord. Further detailed study of the library was performed since the sequence analysis of the fragment showed that it encoded a novel serine protease. The longest clone of the 3′-RACE products obtained using primer 2 and adaptor primer contained 1213 bp including a poly(A) tract. The sequence of the 5′-RACE product was 1381 bp long, and its 3′-end overlapped with the 5′-end of the 3′-RACE product (Table I), enabling the determination of the apparent full-length cDNA of 2265 bp including the 5′- and 3′-untranslated region for a novel serine protease. The longest open reading frame was 1371 bp, which encoded 457 amino acids. This protein was termed spinesin for spinal cord-enriched trypsin-like protease. A homology search against the GenBank™ data base showed that the human spinesin gene spanned 18.8 kb on chromosome 11 and was composed of 13 exons and 12 introns (Fig. 1). The GT-AG rule for exon-intron boundaries was conserved except for the 5′-donor site of the eighth intron, the sequence of which was GC. The cDNA sequence contained three possible initiation codons at the 5′-end, the third codon conforming best to the Kozak consensus sequence (Fig. 1).

Structure of Human Spinesin Deduced from Nucleotide Sequence—Hydropathy plots (Fig. 2) revealed an apparent hydro-

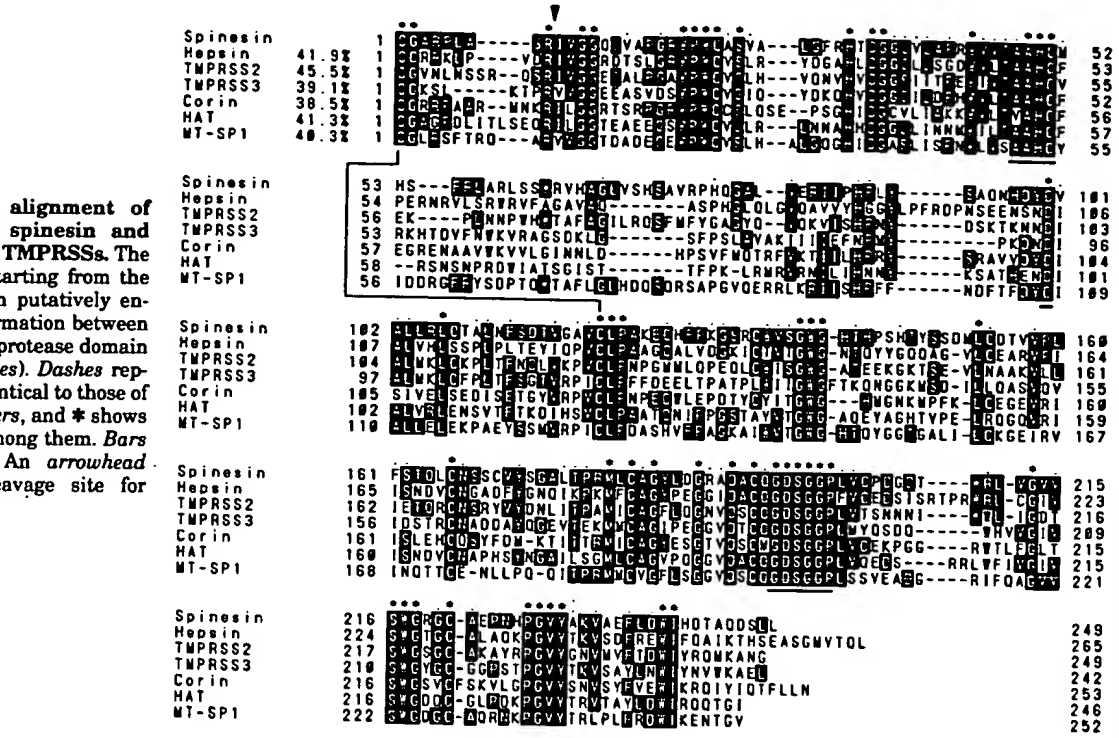


FIG. 3. **Amino acid alignment of protease domains of spinesin and representative human TMPRSSs.** The sequences are aligned starting from the cysteine in the proregion putatively engaged in the disulfide formation between a cysteine residue of the protease domain (shown by *connecting lines*). *Dashes* represent gaps. Residues identical to those of spinesin are in *white letters*, and * shows complete conservation among them. *Bars* show essential triads. An *arrowhead* shows the putative cleavage site for activation.

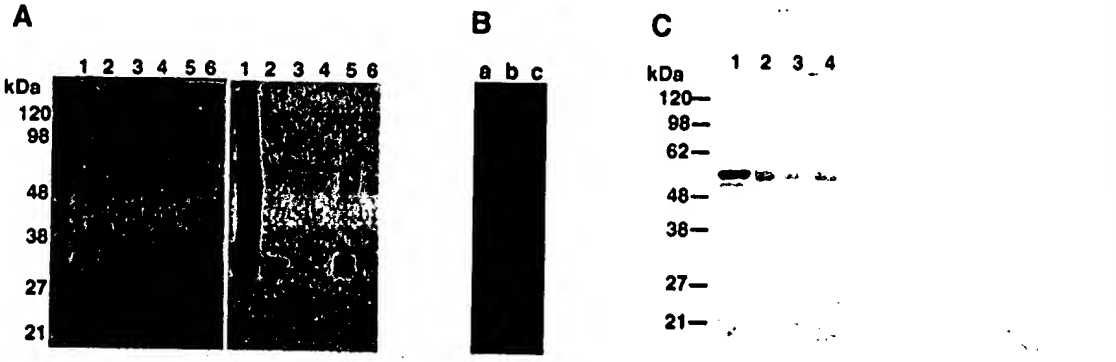


FIG. 4. **Gel electrophoretic analyses of recombinant and natural spinesin.** A fraction obtained from the purification of recombinant chimeric spinesin, H_6DDDDK-(^{218}I–L^{457}), was subjected to 12.5% SDS-PAGE under reducing conditions. *A*, *left* and *right panels* show Coomassie Brilliant Blue staining and Western blot using anti-spinesin B, respectively. *Lane 1*, purified inclusion body; *lane 2*, fraction passed through a Talon column; *lanes 3* and *4*, consecutive fractions of 5 mM imidazole buffer wash; *lanes 5* and *6*, consecutive fractions of 100 mM imidazole buffer eluate. *B*, enzyme fractions electrophoresed through gelatin-polyacrylamide gel as described under "Experimental Procedures." *Lane a*, nontreated/Talon-purified chimeric spinesin; *lane b*, 0.05 unit of enterokinase in 20 mM Tris-HCl, pH 8.0, containing 0.2 M NaCl; *lane c*, enterokinase-treated (0.05 unit) chimeric spinesin. *C*, Western blot analysis of brain homogenate and CSFs. *Lane 1*, brain extract; *lanes 2–4*, CSF samples from three non-CNS diseased patients.

phobic region at Ala^{50}–Leu^{72}, suggesting it to be a transmembrane portion. Both ends of the transmembrane sequence are flanked by a Cys residue that might form a disulfide bond with another Cys residue on each side of the membrane. Five putative *N*-glycosylation sites exist on the sequence C-terminal to the transmembrane portion, suggesting that the molecule is a type II transmembrane glycoprotein (Figs. 1 and 2). Accordingly, the N-terminal cytoplasmic domain and the N-terminal sequence of the stem region between the transmembrane and scavenger receptor-like domain (see below) each carry only two Cys residues that might form a disulfide bridge on each side of the membrane, *i.e.* Cys^{41}–Cys^{49} and Cys^{73}–Cys^{93}. A serine protease domain was located at Ile^{218}–Leu^{457} in the C-terminal half of the molecule and contained the HDS (His, Asp, Ser) triad essential for catalytic activity of a serine protease (Figs. 1, 2, and 3). The stem region connecting the transmembrane and catalytic domains spans from Cys^{73} to Arg^{217} and carries a scavenger receptor-like domain at Val^{110}–Gly^{152} that contains two cysteines that probably form a disulfide bond, Cys^{135}–Cys^{148} (15). The disulfide bridge linking pro- and catalytic domains seems to be formed between Cys^{209} and Cys^{328}, which are conserved among TMPRSSs (Fig. 3). (15). A homology search using the protease domain of spinesin showed that it shares 45.5, 41.9, 41.3, 40.3, 39.1, and 38.5% amino acids with human TMPRSS2, hepsin, HAT, MT-SP1, TMPRSS3, and corin, respectively (Fig. 3). Nine of 10 cysteine residues in the mature enzyme domain of spinesin were well conserved among other TMPRSSs. A putative cleavage site for processing to generate a mature form is tentatively assigned between Arg^{217} and Ile^{218}, which is in the highly conserved activation motif of the serine protease (15). Of the five putative *N*-glycosylation sites, three are in the stem region, and two are in the mature enzyme region (Fig. 2).

Enzyme Characteristics of Recombinant Spinesin Produced in E. coli—To demonstrate that the putative serine protease domain of spinesin has enzymatic activity, a chimeric protein in which Ile^{218}–Leu^{457} of spinesin was fused downstream of H_6DDDDK was expressed in *E. coli*. The products were purified from extensively washed and solubilized inclusion bodies using Talon chelate column chromatography. SDS-PAGE of the

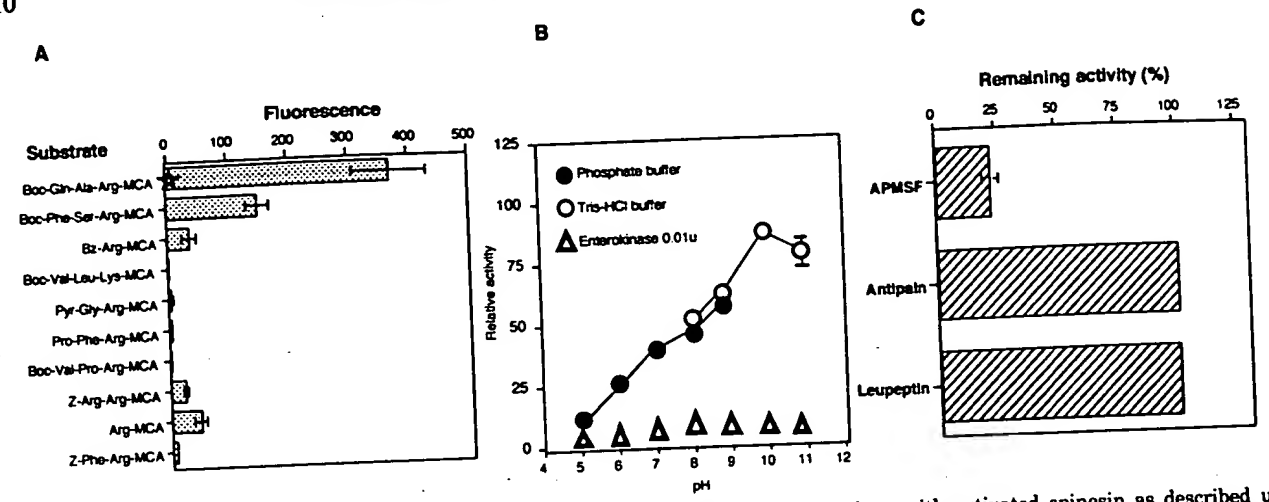


FIG. 5. **Enzymatic characteristics of the recombinant spinesin.** Experiments were done with activated spinesin as described under "Experimental Procedures." *A*, activity against synthetic substrates. Spinesin was activated with an enterokinase column. ☆ shows the activity by the buffer passed through the enterokinase column. *B*, effect of pH on spinesin activity analyzed using 0.1 M phosphate buffer containing 0.2 M NaCl (●) and 0.1 M Tris-HCl buffer containing 0.2 M NaCl (○). △ shows the activity of enterokinase (0.01 unit). *C*, effect of enzyme inhibitors on spinesin activity. *APMSF*, *p*-amidinophenylmethanesulfonyl fluoride hydrochloride; *u*, unit.

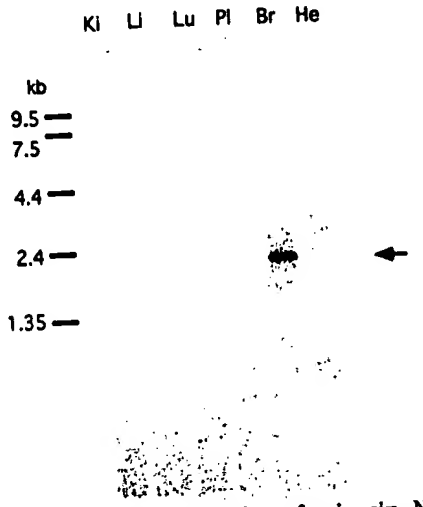


FIG. 6. **Northern blot hybridization of spinesin.** Northern blot hybridization against mRNA from multiple human tissues was carried out as described under "Experimental Procedures." RNA was from kidney (*Ki*), liver (*Li*), lung (*Lu*), placenta (*Pl*), brain (*Br*), and heart (*He*). An *arrow* indicates the 2.3-kbp band.

purified protein showed a single band (Fig. 4*A*, *left*, *lane 5*) of 30 kDa, which was immunoreactive with the anti-human spinesin B in Western blot analysis (Fig. 4*A*, *right*, *lane 5*). The purified recombinant spinesin was activated by treatment with an immobilized enterokinase column to remove H_6DDDDK. The activated spinesin cleaved the synthetic trypsin substrate Boc-Gln-Ala-Arg-MCA (Fig. 5*A*). Using this substrate, the pH optimum was estimated to be about pH 10 (Fig. 5*B*). *p*-Amidinophenylmethanesulfonyl fluoride hydrochloride inhibited spinesin activity by more than 75% at 1 μM, but antipain and leupeptin showed no inhibitory effect at 1 mM (Fig. 5*C*).

As is shown in Fig. 4*B*, the recombinant spinesin gained an ability to cleave gelatin after treatment with 0.05 unit of enterokinase, while neither purified recombinant spinesin before activation or 0.05 unit of enterokinase itself cleaved gelatin. Interestingly casein is not cleaved by the activated recombinant spinesin (data not shown).

Localization of Spinesin in the Human CNS—First, Northern blot analysis was performed to see which human tissues produce spinesin mRNA using a commercially available RNA blot. A clear band was observed at 2.3 kbp in brain but not in kidney, liver, lung, placenta, or heart suggesting a specific expression of spinesin in the CNS (Fig. 6).

The presence of spinesin was verified by Western blot analysis of human brain homogenate, which showed a protein band at 52 kDa that was detectable with anti-human spinesin A anti-serum (Fig. 4*C*). The size suggests that it is a full-length spinesin with possibly five *N*-glycosylated sugar chains. No other distinct bands were apparent using anti-human spinesin B suggesting that the 52-kDa protein is the major molecule present in the brain (data not shown). Spinesin was also detected in the CSF. In some cases, a more rapidly migrating band at about 50 kDa seems dominant (Fig. 4*C*, *lanes 2* and *4*).

Detailed immunohistochemical analysis using anti-human spinesin A showed that at the anterior horn of the spinal cord, neuronal cells and their axons were stained (Fig. 7*A*). The transverse section of the spinal cord revealed many axons to be positively stained (Fig. 7*B*), and among the oligodendrocytes, sporadic staining of the cytoplasm and dendrites was evident (Fig. 7*C*). In addition, the synapses on motor neurons were also stained (Fig. 7*D*). Neuronal cells of the substantia nigra and oculomotor nerve were also strongly stained as well as their axons (data not shown).

DISCUSSION

We have cloned a cDNA encoding a serine protease, designated spinesin or TMPRSS5, of a tentative size of 457 amino acid residues from a human spinal cord cDNA library. It apparently belongs to the TMPRSS family having an N-terminal cytoplasmic domain, a transmembrane domain, a scavenger receptor-like domain in the stem region, and a protease domain. As far as we know, this is the first report of a TMPRSS cloned and identified from CNS. A homology search against GenBank™ showed that the spinesin gene is located on chromosome 11q23, where TMPRSS3 gene is also located; human MT-SP1 is located on the same chromosome. On the other hand, human enterokinase and TMPRSS2 are located at 21q21 and 21q22.3, respectively. Interestingly the gene structure of spinesin is highly similar to that of TMPRSS3 (11). Both are composed of 13 exons spanning 21–24 kb. The protease domain is encoded on exons 8–13, and the transmembrane domain is encoded on the exons 3 and 4. The stem region is encoded on exons 4–8. The gene structure of both human hepsin and TMPRSS2 is also similar to that of spinesin, so it seems possible that the TMPRSS gene family

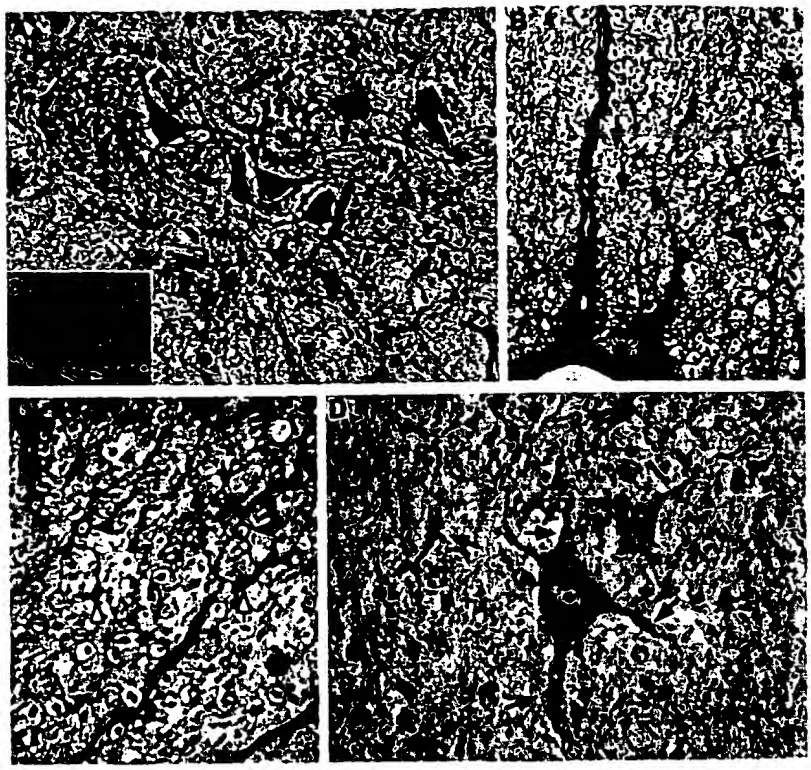


FIG. 7. **Immunohistochemistry of spinesin in human spinal cord.** Experiments were performed as described under "Experimental Procedures." *A*, anterior horn of the spinal cord. Representative positive neuronal cells (*arrow*) and their axons (*arrowhead*) are indicated. *a*, no positive signals in a similar region immunostained using normal rabbit serum. *B*, transverse section of the spinal cord. Representative positive myelinated axons (*arrow*) and oligodendrocytes (*arrowhead*) are shown. *C*, magnified view of a transverse section of the spinal cord. The sporadic staining of cell bodies (*white arrow*) and dendrites of oligodendrocytes (*black arrowhead*) is demonstrated. *D*, magnified view of the anterior horn. *Arrows* indicate spot-like staining of synapses in contact with motoneurons.

share a common ancestor, although they are located on independent chromosomes.

Like TMPRSS2 and TMPRSS3, spinesin has a rather short sequence. Although the longest open reading frame of human spinesin was estimated to be 457 amino acids, the first initiation codon might be a pseudo-codon. It is located on the first exon that is noncoding in the case of TMPRSS3. Two other possible initiation codons present within 30 bp downstream of the first codon are located on the second exon where the initiation codon for TMPRSS3 resides. The third initiation codon best matches the Kozak consensus sequence and may thus be the actual initiation codon. This would make the length of the spinesin 447 amino acid residues. However, further analysis is required to elucidate the real initiation site(s).

Spinesin has a rather simple domain structure carrying only a scavenger receptor-like domain in the stem region like TMPRSS2 and TMPRSS3. The latter two have, in addition, a low density lipoprotein receptor class A domain.

Whereas the other domains are considerably different, the serine protease domains share a high degree of amino acid sequence identity among TMPRSSs. In the protease domain, spinesin showed the highest degree of similarity with TMPRSS2, sharing 45.5% identity of amino acids, followed by TMPRSS3 at 39.1%. [illegible] These enzymes are activated by some as yet unidentified processing enzyme(s) that cleaves the amide bond C-terminal to a lysine or arginine residue in a highly conserved activation site. By analogy with hepsin, we assigned the cleavage site between Arg^{217}–Ile^{218}, which precedes Ile^{218}-Val-Gly-Gly, the well conserved N-terminal sequence among other activated TMPRSSs (Fig. 3). Actually when a chimeric recombinant spinesin produced in *E. coli* was purified and cleaved at the site corresponding to Arg^{217}–Ile^{218} of spinesin (Fig. 1), it showed enzymatic activity against synthetic substrates for trypsin, kallikrein, and plasminogen activator. This result is in accordance with the presence of an aspartate six residues before the catalytic serine (Fig. 3), which would be positioned at the bottom of the S1 substrate binding pocket like in other TMPRSS (1) (Fig. 3).

By analogy also, we predict that the cleaved catalytic domain is linked with the C-terminal side of the stem region by a disulfide bond formed between Cys^{209} and Cys^{328} (Figs. 1 and 3). However, Western blot analysis using anti-spinesin A and B on the brain homogenate revealed the presence of a single major 52 kDa band that is far bigger than the predicted catalytic domain of 240 amino acid residues even if it was *N*-glycosylated at two sites. Whether a smaller active form of the enzyme is present at a level below the detection limit and/or uncleaved 52-kDa spinesin has enzymatic activity remains to be seen. The mechanism underlying the production of 50-kDa spinesin in the CSF and whether the enzyme has activity are also left for future studies.

Immunohistochemical analysis along with Northern blot analysis showed clearly that spinesin is located in the CNS. The neuronal cells and their axons at the anterior horn of the spinal cord were clearly immunopositive. Spinesin was stained in the substantia nigra, oculomotor nucleus, and temporal lobe (data not shown). Furthermore, spinesin was demonstrated at the synapses of the spinal cord. From these results, it seems that spinesin produced in the neuronal cytoplasm may be transported along the axons to the synapses at the anterior horn of the spinal cord. We predict that spinesin is present in the presynaptic regions. As shown in Fig. 7, *B* and *C*, the oligodendrocytes were also stained. The transverse section of the spinal cord clearly demonstrated spinesin in both neuronal axons and oligodendrocytes. The physiological roles of spinesin in the synapses and oligodendrocytes could be different naturally, and further analysis is required to elucidate spinesin functions in different cell types. It should be noted here that, among TMPRSSs, only TMPRSS2 was reported to be present in the brain, having been detected at the mRNA level using human RNA master blot (8), but further examinations such as experiments on the cellular localization of the protein are needed.

The proteolytic activities of membrane-anchored proteins such as membrane-type metalloproteinases and ADAM (a disintegrin-like and metalloproteinase) may play roles in activating events that take place on the cell surface. These enzymes

also may interact with extracellular matrices and proteins on adjacent cells. The enzymatic activity of a few TMPRSSs has been demonstrated. Gelatin, fibrinogen, fibronectin, and laminin are cleaved by TMPRSSs (15). Corin is a processing enzyme of proatrial natriuretic peptide (16), and matriptase processes hepatocyte growth factor as an activator (17). Activated spinesin mainly cleaved trypsin substrates among synthetic forms and cleaved gelatin but not casein.

In summary, we have cloned spinesin/TMPRSS5, a protein that encodes 457 amino acids including a cytoplasmic domain, transmembrane domain, a scavenger receptor-like domain, and a serine protease domain, from human spinal cord mRNA. Spinesin is dominantly expressed at synapses. We predict that axonal spinesin is transported to synaptic junctions for cleavage of protein(s) in the presynaptic regions and that the spinesin dominantly expressed in some oligodendrocytes may activate or inactivate other proteins on the cell surface. We are continuing our efforts to elucidate the biological and pathophysiological functions of spinesin including identifying physiological substrates, interacting molecules, and the exact localization of the molecule in the body including CSF.

REFERENCES

1. La Vallie, E. R., Rehemtulla, A., Racie, L. A., DiBlasio, E. A., Ferenz, C., Grant, K. L., Light, A., and McCoy, J. M. (1993) *J. Biol. Chem.* **268**, 23311–23317
2. Leytus, S. P., Loeb, K. R., Hagen, F. S., Kurachi, K., and Davie, E. W. (1988) *Biochemistry* **27**, 1067–1074
3. Yamaoka, K., Masuda, K., Ogawa, H., Takagi, K., Umemoto, N., and Yasuoka, S. (1998) *J. Biol. Chem.* **273**, 11895–11901
4. Yan, W., Sheng, N., Seto, M., Morser, J., and Wu, Q. (1999) *J. Biol. Chem.* **274**, 14926–14935
5. Hooper, J. D., Scarman, A. L., Clarke, B. E., Normyle, J. F., and Antalis, T. M. (2000) *Eur. J. Biochem.* **267**, 6931–6937
6. Kim, M. G., Chon, C., Lyu, M. S., Cho, E. G., Park, D., Kozak, C., and Schwartz, R. H. (1999) *Immunogenetics* **49**, 420–428
7. Lin, C. Y., Anders, J., Johnson, M., Sang, Q. A., and Dickson, R. B. (1999) *J. Biol. Chem.* **274**, 18231–18236
8. Paoloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C., and Antonarakis, S. E. (1997) *Genomics* **44**, 309–320
9. Lin, B., Ferguson, C., White, J. T., Wang, S., Vessella, R., True, L. D., Hood, L., and Nelson, P. S. (1999) *Cancer Res.* **59**, 4180–4184
10. Wallrapp, C., Hahnel, S., Muller-Pillasch, F., Burghardt, B., Iwamura, T., Ruthenburger, M., Lerch, M. M., Adler, G., and Gress, T. M. (2000) *Cancer Res.* **60**, 2602–2606
11. Scott, H. S., Kudoh, J., Wattenhofer, M., Shibuya, K., Berry, A., Chrast, R., Guipponi, M., Wang, J., Kawasaki, K., Asakawa, S., Minoshima, S., Younus, F., Mehdi, S. Q., Radhakrishna, U., Papasavvas, M. P., Gehrig, C., Rossier, C., Korostishevsky, M., Gal, A., Shimizu, N., Bonne-Tamir, B., and Antonarakis, S. E. (2001) *Nat. Genet.* **27**, 59–63
12. Goldstein, L. A., Ghersi, G., Pineiro-Sanchez, M. L., Salamone, M., Yeh, Y., Flessate, D., and Chen, W. T. (1997) *Biochim. Biophys. Acta* **1361**, 11–19
13. Underwood, L. J., Shigemasa, K., Tanimoto, H., Beard, J. B., Schneider, E. N., Wang, Y., Parmley, T. H., and O'Brien, T. J. (2000) *Biochim. Biophys. Acta* **1502**, 337–350
14. Tanimoto, H., Underwood, L. J., Wang, Y., Shigemasa, K., Parmley, T. H, and O'Brien, T. J. (2001) *Tumor Biol.* **22**, 104–114
15. Hooper, J. D., Clements, J. A., Quigley, J. P., and Antalis, T. M. (2001) *J. Biol. Chem.* **276**, 857–860
16. Yan, W., Wu, F., Morser, J., and Wu, Q. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8525–8529
17. Lee, S.-L., Dickson, R. B., and Lin, C.-Y. (2000) *J. Biol. Chem.* **275**, 36720–36725
18. Kitamoto, Y., Veile, R. A., Donis-Keller, H., and Sadler, J. E. (1995) *Biochemistry* **34**, 4562–4568
19. Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, Y., Nakazato, H., and Yamaguchi, N. (1997) *Biochim. Biophys. Acta* **1352**, 11–14
20. Mitsui, S., Yamada, T., Okui, A., Kominami, K., Uemura, H., and Yamaguchi, N. (2000) *Biochem. Biophys. Res. Commun.* **272**, 205–211
21. Mitsui, S., Okui, A., Kominami, K., Uemura, H., and Yamaguchi, N. (2000) *Biochim. Biophys. Acta* **1494**, 206–210
22. Yamamura, Y., Yamashiro, K., Tsuruoka, N., Nakazato, H., Tsujimura, A., and Yamaguchi, N. (1997) *Biochem. Biophys. Res. Commun.* **239**, 386–392
23. Iijima, N., Tanaka, M., Mitsui, S., Yamamura, Y., Yamaguchi, N., and Ibata, Y. (1999) *Mol. Brain Res.* **66**, 141–149
24. Mitsui, S., Tsuruoka, N., Yamashiro, K., Nakazato, H., and Yamaguchi, N. (1999) *Eur. J. Biochem.* **260**, 627–634
25. Hoop, T. P., and Woods, K. R. (1983) *Mol. Immunol.* **20**, 483–489

EXHIBIT C

October 15, 1992 VOLUME 267 NUMBER 29

ISSN 0021-9258
JBCHA3 267(29) 20525–21272 (1992)

THE Journal of Biological Chemistry

Published by the American Society for Biochemistry and Molecular Biology

FOUNDED BY CHRISTIAN A HERTER
AND SUSTAINED IN PART BY THE CHRISTIAN A HERTER MEMORIAL FUND

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 29, Issue of October 15, pp. 20831-20834, 1992
Printed in U.S.A.

The Cloned Platelet Thrombin Receptor Couples to at Least Two Distinct Effectors to Stimulate Phosphoinositide Hydrolysis and Inhibit Adenylyl Cyclase*

(Received for publication, February 26, 1992)

David T. Hung‡§, Yung H. Wong¶, Thien-Khai H. Vu§, and Shaun R. Coughlin§‖**

From the ‡Department of Laboratory Medicine, the §Cardiovascular Research Institute, the Department of ¶Pharmacology, and the ‖Department of Medicine, University of California, San Francisco, California 94143

Thrombin both stimulates phosphoinositide hydrolysis and inhibits adenylyl cyclase in a variety of cell types. Whether the cloned human platelet thrombin receptor accounts for both of these signaling events is unknown. We report that thrombin receptor agonist peptide causes both phosphoinositide hydrolysis and inhibition of adenylyl cyclase in naturally thrombin-responsive CCL-39 cells. To exclude the possibility that the agonist peptide or thrombin itself may activate these pathways via distinct receptors and to circumvent a lack of suitable thrombin receptor-null cells, we utilized a designed "enterokinase receptor," a thrombin receptor with its thrombin cleavage recognition sequence LDPR replaced by DDDDK, the enterokinase cleavage recognition sequence. Transfection of enterokinase-unresponsive cells with this construct conferred both enterokinase-sensitive phosphoinositide hydrolysis and inhibition of adenylyl cyclase. The phosphoinositide hydrolysis response was largely insensitive to pertussis toxin, whereas the adenylyl cyclase response was completely blocked by pertussis toxin. These data show that the cloned thrombin receptor can effect both phosphoinositide hydrolysis and inhibition of adenylyl cyclase via at least two distinct effectors, most likely G_q-like and G_i-like G-proteins.

Thrombin causes both phosphoinositide hydrolysis and inhibition of adenylyl cyclase in a variety of responsive cells (1–8). Activation of these important second messenger pathways presumably mediates the host of cellular events caused by thrombin, events important in hemostasis and in inflammatory and proliferative responses to vascular injury (9). We recently isolated a cDNA encoding a functional thrombin receptor (10). Whether the cloned receptor accounts for activation of both phosphoinositide hydrolysis and inhibition of cyclase is unknown, and the G-protein(s) which couple to the cloned thrombin receptor have not been identified.

The cloned thrombin receptor is a member of the seven transmembrane span receptor family. The thrombin receptor's 100-residue amino-terminal extension bears a thrombin cleavage site (LDPR/SFLL with/representing the point of cleavage). Structure-function studies (10–12) have shown that thrombin cleaves the receptor at this site, unmasking a new receptor amino terminus beginning with the sequence SFLL. . .; this new amino terminus then functions as a tethered peptide ligand, binding to an as yet undefined receptor domain to effect receptor activation (10). A synthetic peptide representing the new amino terminus unmasked upon receptor cleavage by thrombin is a full agonist for activation of the cloned receptor, bypassing the requirement for proteolysis for receptor activation (10). The ability of the agonist peptide to mimic known cellular actions of thrombin has been utilized to implicate a role for the cloned receptor in specific thrombin responses (10, 13).

We report here that the human platelet thrombin receptor agonist peptide causes both phosphoinositide hydrolysis and inhibition of adenylyl cyclase in naturally thrombin-responsive CCL-39 fibroblasts. Pouyssegur and colleagues (14) have also observed activation of these two important second messenger pathways in CCL-39 fibroblasts using a similar hamster thrombin receptor agonist peptide. These findings suggest but do not prove that the cloned thrombin receptor may mediate both phosphoinositide hydrolysis and inhibition of adenylyl cyclase. The possibility that the agonist peptide might act at two distinct receptors to elicit these two well described thrombin-induced cellular signaling events has not been addressed.

To exclude the possibility that the agonist peptide or thrombin itself activates phosphoinositide hydrolysis and inhibits adenylyl cyclase via distinct receptors and to circumvent a lack of a suitable thrombin receptor null cell, we utilized an artificial "enterokinase receptor," a thrombin receptor in which the thrombin cleavage recognition site LDPR was replaced with DDDDK, the enterokinase cleavage recognition sequence (11). Most cell lines tested thus far have been null for enterokinase responses. Moreover, after cleavage-induced activation, the sequence of the enterokinase receptor becomes identical to that of the activated native thrombin receptor (11). Post-activation signaling by the two receptors should be identical.

Transfection of cells with enterokinase receptor cDNA conferred both enterokinase-inducible phosphoinositide hydrolysis and inhibition of adenylyl cyclase. The phosphoinositide hydrolysis response was largely pertussis toxin-insensitive, whereas the adenylyl cyclase response was completely blocked by pertussis toxin. These data show that the cloned thrombin receptor can effect both phosphoinositide hydrolysis and inhibition of adenylyl cyclase via at least two distinct effectors.

* This work was supported in part by Physician-Scientist Award NIH HL0248-01 (to D. T. H.), NIH GM27800 (to Y. H. W.), NIH HL44907 and by Tobacco-related Disease Research Program of the University of California Grants 2RT19 and HL43801 (to S. R. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** This work was done during the tenure of an award from the American Heart Association and SmithKline Beecham. To whom correspondence should be addressed: Box 0524, University of California, San Francisco, CA 94143.

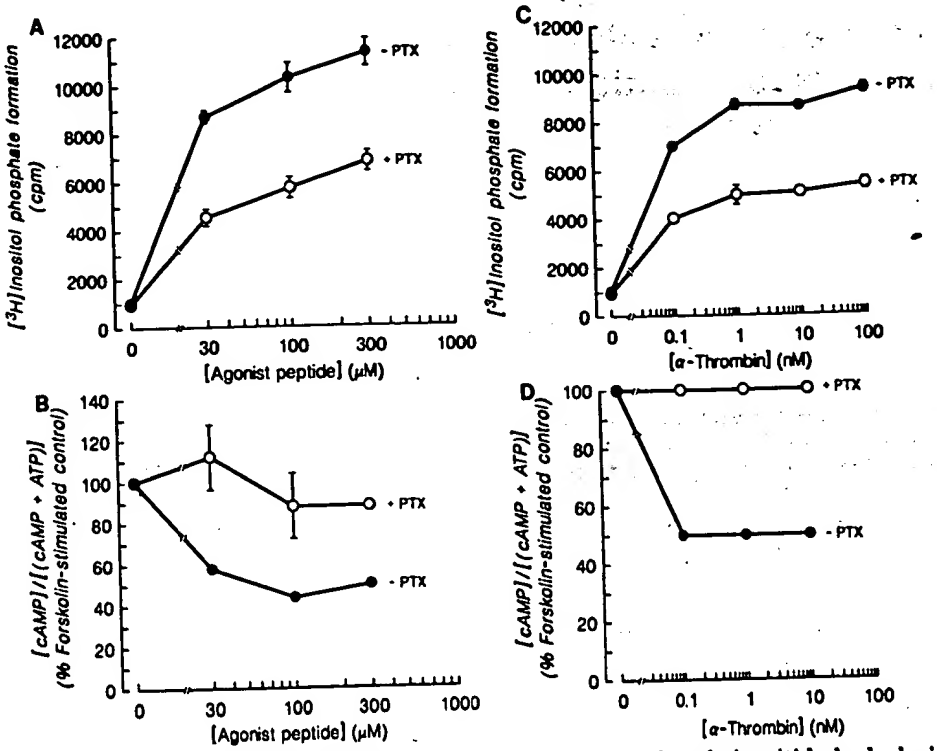


FIG. 1. **Thrombin receptor agonist peptide recapitulates thrombin-induced phosphoinositide hydrolysis and inhibition of adenylyl cyclase in CCL-39 fibroblasts.** *A* and *C*, peptide- or thrombin-induced phosphoinositide hydrolysis. Quiesced CCL-39 fibroblasts loaded with *myo*-[^{3}H]inositol (see "Experimental Procedures") were treated in the presence of 20 mM LiCl for 15 min at 37 °C with thrombin receptor agonist peptide (*A*) or thrombin (*C*) at the concentrations indicated. Total [^{3}H]inositol phosphates were collected and quantitated as described under "Experimental Procedures." *B* and *D*, peptide or thrombin-induced inhibition of adenylyl cyclase. Quiesced CCL-39 fibroblasts loaded with [^{3}H]adenine (see "Experimental Procedures") were treated with 1 mM IBMX and 5 μM forskolin in the absence of agonists or in the presence of thrombin receptor agonist peptide (*B*) or thrombin (*D*) at the concentrations indicated. Intracellular cAMP accumulation was estimated by determining the ratios of radiolabeled cAMP to total ATP plus ADP as described under "Experimental Procedures." For all assays, cells were preincubated at 37 °C for 5 h in the presence (*open circles*) or absence (*closed circles*) of pertussis toxin at a final concentration of 100 ng/ml. Each point represents the mean of triplicate determinations with standard deviations as shown. These results are representative of three independent experiments.

EXPERIMENTAL PROCEDURES

Materials—Purified human α-thrombin was a generous gift from Dr. John W. Fenton II, Albany Medical College of Union University, Albany, NY. Thrombin receptor agonist peptide SFLLRNPNDKYEPF and "scrambled" peptide FSLLRNPNDKYEPF were synthesized by University of California San Francisco's Biomolecular Resource Center. Peptides were purified by reversed phase HPLC before use. Purified bovine enterokinase was obtained from Biozyme Corporation, San Diego, CA. [^{3}H]Adenine was obtained from Du Pont-New England Nuclear, 1-methyl-3-isobutylxanthine (IBMX),[1] forskolin, trichloroacetic acid, cAMP, and ATP were obtained from Sigma. All cell culture plasticware was obtained from Falcon. DMEM and G418 were obtained from GIBCO. *Bordatella pertussis* toxin islet-activating protein was obtained from List Biological Laboratories, Campbell, CA.

Receptor-expressing Cell Lines—Stably transfected Rat 1 cells expressing the cloned wild type human platelet thrombin receptor and mutant enterokinase-activated thrombin receptor were prepared as described previously (8, 11).

Phosphoinositide Hydrolysis Assays—These assays were performed as described previously (8).

cAMP Accumulation Assays—Inhibition of adenylyl cyclase was assessed in cells stimulated with forskolin. Untransfected Rat 1 cells and Rat 1 cells expressing the native human platelet thrombin receptor or the mutant enterokinase receptor were grown in 24-well plates (Falcon) in 1 ml of Hepes-buffered DMEM (containing 100 units/ml penicillin, 100 μg/ml streptomycin, 25 mM Hepes buffer, pH 7.4) supplemented with 10% fetal calf serum, and upon reaching confluence, were rinsed with 1 ml of Hepes-buffered DMEM and then incubated with 2 μCi/ml [^{3}H]adenine in the same medium for 16–20 h at 37 °C. All wells were then aspirated and rinsed with 1 ml of Hepes-buffered DMEM containing 1 mM IBMX followed by incubation with 200 μl of Hepes-buffered DMEM containing 1 mM IBMX and 5 μM forskolin in the presence or absence of the agonists thrombin, thrombin receptor agonist peptide, or enterokinase for 30 min at 37 °C. All wells were then aspirated and fixed in 5% trichloroacetic acid with 1 mM cAMP and 1 mM ATP for at least 30 min at 4 °C. Intracellular [^{3}H]cAMP accumulation was estimated by determining the ratios of radiolabeled cAMP to total ATP plus ADP as described previously (15, 16).

Pertussis Toxin Treatment—In all experiments involving the use of *B. pertussis* toxin islet-activating protein, cells were incubated at 37 °C for 5 h before agonist addition in the presence of pertussis toxin at a final concentration of 100 ng/ml as described previously (8). Treatment of cells with this concentration of pertussis toxin was shown to remove all available pertussis substrate from cell membranes, as detected by subsequent ADP-ribosylation of G-protein subunits *in vitro* (8).

RESULTS AND DISCUSSION

Thrombin receptor agonist peptide both stimulated phosphoinositide hydrolysis and inhibited adenylyl cyclase in CCL-39 cells (Fig. 1, *A* and *B*). The maximum responses elicited by saturating concentrations of agonist peptide (100 μM) or thrombin (1 nM) were similar (Fig. 1, *C* and *D*). The relative potencies of agonist peptide and thrombin in this system were similar to those observed in oocytes (10). The higher potency of thrombin compared with the agonist peptide is consistent with the current model of thrombin receptor activation in which thrombin acts enzymatically to generate a tethered agonist ligand, whereas the synthetic agonist pep-

[1] The abbreviations used are: IBMX, 1-methyl-3-isobutylxanthine; DMEM, Dulbecco's modified Eagles's medium.

tide is not tethered and acts by a nonenzymatic occupancy-dependent mechanism.

Both agonist peptide- and thrombin-induced inhibition of adenylyl cyclase in CCL-39 fibroblasts were completely pertussis toxin-sensitive (Fig. 1, *B* and *D*). By contrast, agonist peptide- and thrombin-induced stimulation of phosphoinositide hydrolysis was only partially pertussis toxin-sensitive (Fig. 1, *A* and *C*). This differential pertussis toxin sensitivity suggests that distinct effectors mediate these two thrombin responses. Gq is a recently described G-protein which couples to the β1 isozyme of phospholipase C and lacks a pertussis toxin site (17). Thrombin-induced phosphoinositide hydrolysis may thus be mediated by a G_q-like G-protein. G_i mediates inhibition of adenylyl cyclase in a completely pertussis toxin-sensitive manner (18). A G_i-like G-protein presumably mediates inhibition of adenylyl cyclase by thrombin.

The results shown in Fig. 1 suggested that the endogenous thrombin receptor in CCL-39 cells and presumably the cloned thrombin receptor coupled to both inhibition of adenylyl cyclase and stimulation of phosphoinositide hydrolysis via at least two distinct effectors, probably G_i- and G_q-like G-proteins. Alternatively, these results were consistent with the agonist peptide or thrombin itself acting at two distinct receptors to activate these two second messenger pathways. To distinguish between these alternatives, we set out to perform transfection experiments to determine whether the single molecule encoded by the cloned thrombin receptor cDNA could couple to both phosphoinositide hydrolysis and inhibition of cyclase, but were hampered by lack of a suitable null cell. We had shown previously that CV-1 cells had a minimal endogenous phosphoinositide response to thrombin and that transfection of these cells with the thrombin receptor cDNA conferred robust thrombin-sensitive phosphoinositide hydrolysis (8). When we attempted to utilize these cells to address whether the cloned thrombin receptor might also couple to adenylyl cyclase responses, we noted thrombin-induced increases in cAMP levels in the untransfected parent cells. This observation may reflect efficient coupling of a small number of endogenous thrombin receptors to G_i and a type II-like adenylyl cyclase (such a coupling mechanism has been proposed for α_2-adrenergic receptor signaling in COS-7 cells (19), which are derived from CV-1 cells). Alternatively, this observation is again consistent with the existence of a second type of thrombin receptor coupled to cyclase responses. Regardless, CV-1 cells were not suitable null cells for our transfection studies. After screening a number of other cell lines for a null phenotype, we noted Rat-1 cells to have only minimal endogenous thrombin-stimulated phosphoinositide hydrolysis. Unfortunately, these cells again had an endogenous cyclase response, a small and variable decrease in cAMP levels in response to thrombin. These cells were thus also unsuitable for transfection studies to address whether the cloned thrombin receptor might couple to both phosphoinositide hydrolysis and inhibition of cyclase.

To circumvent the lack of a suitable null cell line, we utilized an artificial "enterokinase" receptor constructed by replacing the thrombin receptor's thrombin cleavage recognition sequence LDPR with the enterokinase cleavage recognition sequence DDDDK (11). This maneuver resulted in a receptor responsive to the highly specific protease enterokinase, but not to thrombin (11). Most cell lines tested, including Rat-1 cells, have been null for enterokinase responses. Moreover, cleavage of the enterokinase receptor by enterokinase and of the wild type thrombin receptor by thrombin results in activated receptor molecules with identical amino acid sequences. Post-activation signaling should also be identical.

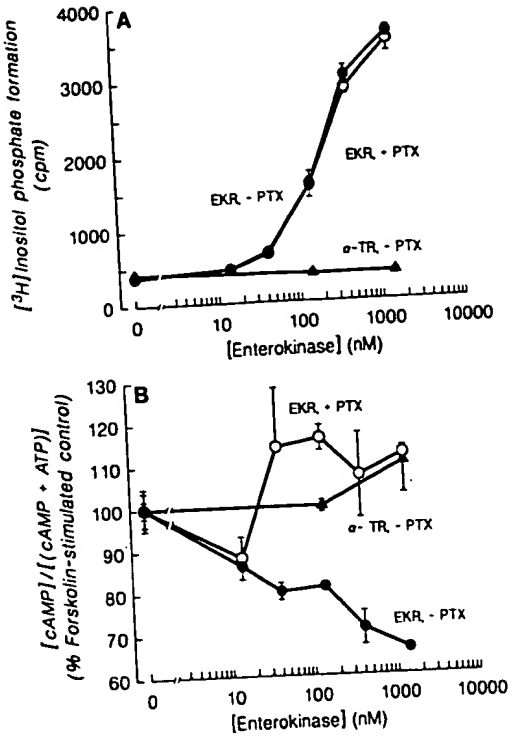


FIG. 2. **Enterokinase causes phosphoinositide hydrolysis and inhibition of adenylyl cyclase in Rat 1 cells expressing the cloned thrombin receptor engineered with an enterokinase cleavage activation sequence.** *A*, quiesced Rat 1 cells stably transfected with cDNA encoding the enterokinase receptor (*circles*) or the native thrombin receptor (*triangles*) were loaded with *myo*-[^{3}H]inositol (see "Experimental Procedures"), incubated for 5 h at 37 °C in the absence (*closed circles*) or presence (*open circles*) of 100 ng/ml pertussis toxin, followed by treatment with 20 mM LiCl and enterokinase at the concentrations indicated for 30 min at 37 °C. Total [^{3}H]inositol phosphates were collected and quantitated as described under "Experimental Procedures." *B*, quiesced Rat 1 cells stably transfected with cDNA encoding the enterokinase receptor (*circles*) or the native thrombin receptor (*triangles*) were loaded with [^{3}H]adenine (see "Experimental Procedures"), incubated for 5 h at 37 °C in the absence (*closed circles*) or presence (*open circles*) of 100 ng/ml pertussis toxin, followed by treatment with 1 mM IBMX, 5 μM forskolin, and enterokinase at the concentrations indicated for 30 min at 37 °C. Intracellular cAMP accumulation was estimated by determining the ratios of radiolabeled cAMP to total ATP plus ADP as described under "Experimental Procedures." For all assays, each point represents the mean of triplicate determinations with standard error of the mean as shown. These results are representative of three independent experiments.

Transfected Rat 1 cells expressing the enterokinase-activated mutant thrombin receptor demonstrated phosphoinositide hydrolysis in response to enterokinase with an EC_{50} of approximately 200 nM, whereas untransfected Rat 1 cells or Rat 1 cells transfected with wild type thrombin receptor cDNA showed no phosphoinositide turnover response to enterokinase (Fig. 2*A*). The magnitude of enterokinase-induced phosphoinositide turnover in Rat 1 cells expressing the "enterokinase thrombin receptor" was similar to the magnitude of thrombin-induced phosphoinositide turnover in Rat 1 cells expressing the wild type thrombin receptor (data not shown). Enterokinase-stimulated phosphoinositide hydrolysis in Rat 1 cells transfected with the enterokinase thrombin receptor was largely pertussis toxin-insensitive (Fig. 2*A*), consistent with the effects of pertussis toxin on thrombin-induced phosphoinositide hydrolysis noted in CV-1 cells transfected with the wild type platelet thrombin receptor and suggesting that the thrombin receptor couples to a G_q-like G-protein (8).

Thrombin receptor coupling to phosphoinositide hydrolysis via as yet undefined pertussis toxin-insensitive pathways cannot, of course, be excluded.

Rat 1 cells transfected with the enterokinase thrombin receptor demonstrated enterokinase-induced inhibition of adenylyl cyclase with an EC_{50} of approximately 100 nM enterokinase. By contrast, untransfected Rat 1 cells or Rat 1 cells expressing wild type thrombin receptor cDNA demonstrated no inhibition of adenylyl cyclase in response to enterokinase (Fig. 2*B*). These data demonstrate that the cloned thrombin receptor can mediate inhibition of adenylyl cyclase as well as stimulation of phosphoinositide hydrolysis. In contrast to phosphoinositide hydrolysis (Fig. 2*A*), enterokinase-induced inhibition of adenylyl cyclase in Rat 1 cells expressing the enterokinase thrombin receptor was completely pertussis toxin-sensitive (Fig. 2*B*). This differential pertussis sensitivity strongly suggests that the thrombin receptor couples to phosphoinositide hydrolysis and inhibition of adenylyl cyclase via at least two different coupling mechanisms. As noted above, G_q- and G_i-like G-proteins are plausible candidates for mediating such coupling.

Although several cloned receptors have been shown to be capable of coupling to more than one G-protein, few appear capable of coupling to both G_i and G_q. The m2 muscarinic receptor and the α_2-adrenergic receptor have been observed to couple to both phosphoinositide hydrolysis and inhibition of adenylyl cyclase via G_q and G_i, respectively, when overexpressed in Chinese hamster ovary (CHO) or COS cells (20, 21). In these systems, these receptors appear to be more efficiently coupled to the regulation of adenylyl cyclase than to phospholipase C. The thrombin receptor's ability to regulate both of these signaling pathways efficiently and in "opposite" directions may account in part for thrombin's unusual effectiveness as an agonist for platelet activation, an event known to be promoted by both activation of phosphoinositide hydrolysis and inhibition of adenylyl cyclase (22).

In summary, these studies provide compelling evidence that the cloned platelet thrombin receptor couples to both phosphoinositide hydrolysis and inhibition of adenylyl cyclase via at least two distinct effectors, likely coupling to both a G_q-like G-protein and a G_i-like G-protein. These data also demonstrate the feasibility of utilizing a designed "artificial" receptor with altered specificity to circumvent the lack of a null cell line to study second messenger cellular signaling pathways.

Acknowledgments—We thank Dr. John Fenton II (Wadsworth Center/Albany Medical College) for his generous gift of α-thrombin and Dr. Henry R. Bourne for valuable discussions.

REFERENCES

1. Paris, S., and Pouyssegur, J. (1986) *EMBO J.* **5(1)**, 55–60
2. Brass, L. F., Laposata, M., Banga, H. S., and Rittenhouse, S. E. (1986) *J. Biol. Chem.* **261**, 16838–16847
3. Jones, L. G., McDonough, P. M., and Brown, J. H. (1989) *Mol. Pharmacol.* **36**, 142–149
4. Brass, L. F., Manning, D. R., Williams, A. G., Woolkalis, M. J., and Poncz, M. (1991) *J. Biol. Chem.* **266**, 958–965
5. Magnaldo, I., Pouyssegur, J., and Paris, S. (1988) *Biochem. J.* **253**, 711–719
6. Crouch, M., and Lapetina, E. G. (1988) *J. Biol. Chem.* **263**, 3363–3371
7. Rasmussen, U. B., Vouret-Craviari, V., Jallat, S., Schlesinger, Y., Pages, G., Pavirani, A., Lecocq, J.-P., Pouyssegur, J., and Van Obberghen-Schilling, E. (1991) *FEBS Lett.* **288**, 123–128
8. Hung, D. T., Vu, T.-K. H., Nelken, N. A., and Coughlin, S. R. (1992) *J. Cell Biol.* **116**, 827–832
9. Shuman, M. A. (1986) *Ann. N. Y. Acad. Sci.* **485**, 228–239
10. Vu, T.-K. H., Hung, D. T., Wheaton, V. I., and Coughlin, S. R. (1991) *Cell* **64**, 1057–1068
11. Vu, T.-K. H., Wheaton, V. I., Hung, D. T., Charo, I., and Coughlin, S. R. (1991) *Nature* **353**, 674–677
12. Liu, L.-W., Vu, T.-K. H., Esmon, C. T., and Coughlin, S. R. (1991) *J. Biol. Chem.* **266**, 16977–16980
13. Huang, R., Sorisky, A., Church, W. R., Simons, E. R., and Rittenhouse, S. E. (1991) *J. Biol. Chem.* **266**, 18435–18438
14. Vouret-Craviari, V., Van Obberghen-Schilling, E., Rasmussen, U. B., Pavirani, A., Lecocq, J.-P., and Pouyssegur, J. (1992) *Mol. Biol. Cell.* **3**, 95–102
15. Wong, Y. H., Federman, A., Pace, A. M., Zachary, I., Evans, T., Pouyssegur, J., and Bourne, H. R. (1991) *Nature* **351**, 63–65
16. Salomon, Y., Londos, C., and Rodbell, M. (1974) *Anal. Biochem.* **58**, 541–548
17. Smrcka, A. V., Hepler, J. R., Brown, K. O., and Sternweis, P. C. (1991) *Science* **251**, 804–807
18. Birnbaumer, L. (1990) *Annu. Rev. Pharmacol. Toxicol.* **30**, 675–705
19. Federman, A. D., Conklin, B. R., Schrader, K. A., Reed, R. R., and Bourne, H. R. (1992) *Nature* **356**, 159–161
20. Peralta, E. G., Ashkenazi, A., Winslow, J. W., Smith, D. H., Ramachandran, J., and Capon, D. J. (1987) *Science* **236**, 600–605
21. Cotecchia, S., Kobilka, B. K., Daniel, K. W., Nolan, R. D., Lapetina, E. Y., Caron, M. G., Lefkowitz, R. J., and Regan, J. W. (1990) *J. Biol. Chem.* **265**, 63–69
22. Kroll, M. H., and Schafer, A. I. (1989) *Blood* **74**, 1181–1195

BIOCHEMICAL JOURNAL

BIJOAK 344(2) 281–623
1999
ISSN 0264-6021

Volume 344
part 2
1 December 1999

Biochem. J. (1999) 344, 331–337 (Printed in Great Britain)

Cloning, expression and characterization of YSA1H, a human adenosine 5′-diphosphosugar pyrophosphatase possessing a MutT motif

Lakhdar GASMI, Jared L. CARTWRIGHT and Alexander G. McLENNAN[1]

School of Biological Sciences, Life Sciences Building, University of Liverpool, P.O. Box 147, Liverpool L69 7ZB, U.K.

The human homologue of the *Saccharomyces cerevisiae* YSA1 protein, YSA1H, has been expressed as a thioredoxin fusion protein in *Escherichia coli*. It is an ADP-sugar pyrophosphatase with similar activities towards ADP-ribose and ADP-mannose. Its activities with ADP-glucose and diadenosine diphosphate were 56% and 20% of that with ADP-ribose respectively, whereas its activity towards other nucleoside 5′-diphosphosugars was typically 2–10%. cADP-ribose was not a substrate. The products of ADP-ribose hydrolysis were AMP and ribose 5-phosphate. K_m and k_{cat} values with ADP-ribose were 60 μM and 5.5 s^{-1} respectively. The optimal activity was at alkaline pH (7.4–9.0) with 2.5–5 mM Mg^{2+} or 100–250 μM Mn^{2+} ions; fluoride was inhibitory, with an IC_{50} of 20 μM. The YSA1H gene, which maps to 10p13–p14, is widely expressed in all human tissues examined, giving a 1.4 kb transcript. The 41.6 kDa fusion protein behaved as an 85 kDa dimer on gel filtration. After cleavage with enterokinase, the 24.4 kDa native protein fragment ran on SDS/PAGE with an apparent molecular mass of 33 kDa. Immunoblot analysis with a polyclonal antibody raised against the recombinant YSA1H revealed the presence of a protein of apparent molecular mass 33 kDa in various human cells, including erythrocytes. The sequence of YSA1H contains a MutT sequence signature motif. A major proposed function of the MutT motif proteins is to eliminate toxic nucleotide metabolites from the cell. Hence the function of YSA1H might be to remove free ADP-ribose arising from NAD^+ and protein-bound poly- and mono-(ADP-ribose) turnover to prevent the occurrence of non-enzymic protein glycation.

Key words: ADP-ribose, nucleotide sugar, nudix hydrolase, protein glycation.

INTRODUCTION

The MutT sequence motif proteins are a recently described family of mostly small, soluble nucleotide pyrophosphatases named after the 16 kDa *Escherichia coli mutT* gene product, a hydrolase with a strong preference for the mutagenic nucleotide 8-oxo-dGTP [1]. They all possess the sequence signature motif GX_5EX_5[UA]XRE[UT]XEEXGU (single-letter codes, where U is a hydrophobic amino acid, and the square brackets indicate that the position could be occupied by any of the amino acids within the brackets) [2]. The *E. coli* genome sequence reveals 11 further proteins with perfect or closely related MutT motifs, several of which have now been cloned and characterized [3]. The substrate specificities of these other proteins varies from quite specific to relatively broad; known substrates include dATP, NADH, nucleoside 5′-diphosphosugars (NDP-sugars) and dinucleoside polyphosphates. Because all these substrates have the general structure of a nucleoside diphosphate linked to another moiety, X, the name 'nudix' hydrolase has been proposed for this family. The function of members of this protein family might be to cleanse the cell of potentially deleterious endogenous nucleotide metabolites and to modulate the accumulation of metabolic intermediates by diverting them into alternative pathways in response to biochemical need [3].

MutT motif proteins are also found in eukaryotes, including humans [4]. Known human proteins include the MutT orthologue MTH1 [5], diadenosine 5′,5‴-P^1,P^4-tetraphosphatase (Ap_4A hydrolase) [6] and diphosphoinositol polyphosphate phosphohydrolase (DIPP) [7]. DIPP is homologous with the recently described Ap_6A hydrolases from budding and fission yeasts and, like them, is remarkable in its ability to hydrolyse the non-nucleotide diphosphoinositol polyphosphates in addition to the structurally unrelated diadenosine polyphosphates [8]. A survey of the GenBank expressed sequence tag (EST) database suggests the existence of at least six further human MutT motif proteins, including a protein encoded in an anti-sense transcript of the basic fibroblast growth factor gene [9], a protein closely related to DIPP (DIPP2) [7], and a homologue of the *Saccharomyces cerevisiae* YSA1 protein. YSA1 has been reported to have ADP-ribose pyrophosphatase activity (EC 3.6.1.13), although no details of this activity have been published [3,10].

ADP-ribose is the most significant ADP-sugar in mammalian cells. It is generated by the turnover of NAD^+ and protein-bound poly- and mono-(ADP-ribose) and, if allowed to accumulate, could be hazardous to the cell by virtue of its ability to modify protein histidine, lysine and cysteine residues by non-enzymic glycation [11–13] and to bind to ATP-activated K^+ channels [14]. The elimination of free ADP-ribose therefore seems to fall within the category of 'housecleaning' functions proposed for the MutT motif family. An ADP-ribose pyrophosphatase has recently been isolated from human erythrocytes [12,15], whereas three cytoplasmic and one mitochondrial ADP-ribose pyrophosphatases in rat liver have been described [16]. A possibly distinct ADP-sugar pyrophosphatase (EC 3.6.1.21) has also been isolated from

Abbreviations used: Ap_4A, diadenosine 5′,5‴-P^1,P^4-tetraphosphate (other diadenosine polyphosphates are abbreviated similarly); DIPP, diphosphoinositol polyphosphate phosphohydrolase; EST, expressed sequence tag; IPTG, isopropyl β-D-thiogalactoside; LB, Luria–Bertani; NDP-sugar, nucleoside 5′-diphosphosugar; NTA, nitrilotriacetate; ORF, open reading frame; Trx, thioredoxin; YSA1H, YSA1 homologue.

[1] To whom correspondence should be addressed (e-mail agmclen@liv.ac.uk).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF155832.

mammalian liver [17]. However, the amino acid sequences of none of these enzymes have been determined. Here we report the cloning, expression and characterization of YSA1H, the human homologue of the yeast YSA1 protein, and show that it is an ADP-sugar pyrophosphatase with its highest activity towards ADP-ribose, ADP-mannose and ADP-glucose.

Since this paper was submitted the Human Genome Nomenclature Committee (HGNC) has approved gene symbols for the human MutT motif proteins. Accordingly theYSA1H gene is designated NUDT5. Full details can be found on http://www.gene.ucl.ac.uk/users/hester/npym.html or by contacting the committee (e-mail nome@galton.ucl.ac.uk.).

EXPERIMENTAL

Materials

Mononucleotides, dinucleoside polyphosphates, NDP-sugars and nucleoside 5′-diphosphoalcohols were from Sigma. [^{32}P]UTP (800 Ci/mmol) was from ICN; NAD^+, NADH, $NADP^+$, NADPH, calf intestinal alkaline phosphatase, yeast inorganic pyrophosphatase, *Eco*RI and *Nco*I were from Boehringer Mannheim; pET 32b(+) was from Novagen; *Pfu* DNA polymerase was from Stratagene; I. M. A. G.E. clone 310860 was obtained from the UK Human Genome Mapping Project Resource Centre.

Cloning of YSA1H cDNA

I.M.A.G.E. clone 310860 contains an insert corresponding to a YSA1H cDNA in the polylinker-modified plasmid pT7T3D. The sequence of the insert was confirmed (see Figure 1) and the insert was then amplified by PCR. The 29-mer oligonucleotide forward and reverse primers 5′-d(TCGTTTGACCATGGAGAGCC AAGAACCAA)-3′ and 5′-d(TAAGTTGAATTCAAAGTCTT-AGTGAAGAA)-3′ were synthesized to provide an *Nco*I restriction site at the start of the amplified insert and an *Eco*RI site at the end. After amplification with *Pfu* DNA polymerase, the DNA was recovered by extraction with phenol/chloroform/3-methylbutan-1-ol (25:24:1, by vol.) and then digested with *Nco*I and *Eco*RI. The digest was gel-purified and the restriction fragment was ligated between the *Nco*I and *Eco*RI sites of the pET32b(+) thioredoxin (Trx) fusion vector (Novagen). The resulting construct, pET-YSA1H, generated a fusion of YSA1H downstream of the cleavable 109-residue Trx fusion protein and His-tag and S-tag sequences. This plasmid was used to transform *E. coli* XL1-Blue cells for propagation.

Protein expression and purification

E. coli strain BL21(DE3) was transformed with pET-YSA1H. A single colony was picked from a Luria–Bertani (LB) agar plate containing 60 μg/ml ampicillin and inoculated into 10 ml LB medium containing 60 μg/ml ampicillin. After overnight growth at 37 °C, the cells were transferred to 1 litre of LB medium containing 60 μg/ml ampicillin and grown at 30 °C to a D_{600} of 0.8. Isopropyl β-D-thiogalactoside (IPTG) was added to 1 mM and the cells were incubated for 4 h. The induced cells were harvested, washed and resuspended in 50 ml of breakage buffer [50 mM Tris/HCl (pH 8.0)/0.1 M NaCl]. The cell suspension was sonicated and the resulting lysate was cleared by centrifugation at 15000 *g* and 4 °C for 10 min. The supernatant was recovered and applied to a 15 mm × 50 mm column of Ni^{2+}/nitrilotriacetate (NTA)-agarose (Sigma) equilibrated with 20 mM Tris/acetate (pH 7.0)/0.3 M NaCl/10 mM 2-mercaptoethanol at a flow rate of 0.5 ml/min. After elution of the unbound proteins, a linear gradient of 0–50 mM histidine in equilibration buffer was applied at a flow rate of 1 ml/min; fractions of 1 ml were collected and analysed by SDS/PAGE. Those containing pure Trx–YSA1H fusion protein were collected and concentrated by ultrafiltration.

Enzyme assays and product identification

Substrates were screened by measuring the P_i released in a coupled assay by co-incubation of substrate with YSA1H protein and either alkaline phosphatase or inorganic pyrophosphatase [16,18]. The standard assay (200 μl) for phosphodiester substrates was incubation for 10 min at 37 °C with 50 mM Tris/HCl (pH 7.5)/5 mM $MgCl_2$/1 mM dithiothreitol/0.3 mM substrate/0.25 μg of Trx–YSA1H fusion protein and 0.5 μg (1 unit) of alkaline phosphatase. Phosphomonoester substrates were assayed as above, except that 0.5 μg (100 m-units) of inorganic pyrophosphatase was used instead of alkaline phosphatase. The P_i released in each case was measured colorimetrically. P_i released from $NADP^+$ and NADPH in control assays without YSA1H was subtracted. Kinetic parameters for ADP-ribose hydrolysis were determined by measuring the P_i released in the alkaline-phosphatase-coupled assay as described. Non-specific alkaline phosphodiesterase (nucleotide pyrophosphatase) activity was assayed with thymidine 5′-monophospho-*p*-nitrophenyl ester as substrate [19].

Reaction products generated from the different substrates were identified by high-performance anion-exchange chromatography. Reaction mixtures containing 50 mM Tris/HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.3 mM substrate and 3 μg of Trx–YSA1H fusion protein were incubated at 37 °C for 30 min in a volume of 200 μl. Samples (100 μl) of the reaction mix were applied to a 1 ml Resource-Q column (Pharmacia) equilibrated with 45 mM ammonium acetate (pH adjusted to 4.6 with H_3PO_4) and eluted with a linear gradient from 0% to 100% (v/v) 0.5 M NaH_2PO_4 (adjusted to pH 2.7 with acetic acid) for 10 min at a flow rate of 1 ml/min [15]. Elution was monitored at 260 nm and peaks were identified with the aid of standards.

Northern and dot-blot analyses

A ^{32}P-labelled anti-sense RNA to the YSA1H cDNA was generated from I.M.A.G.E. clone 310860 by using an *in vitro* transcription kit (Ambion), [^{32}P]UTP and T3 RNA polymerase in accordance with the manufacturer's instructions. A Multiple Choice™ Northern blot (OriGene) containing poly(A)$^+$-selected RNA from various human tissues (2 μg per lane) was probed overnight with the RNA probe (10^6 c.p.m./ml) at 64 °C. The blot was washed twice for 15 min each with SSC (0.15 M NaCl/0.015 M sodium citrate) containing 0.5% SDS at 64 °C. The bound ^{32}P-labelled probe was detected with a PhosphorImager (Bio-Rad). The blot was then stripped for re-use by using the Strip-EZ solutions provided with the transcription kit. A Human RNA Master Blot™ dot-blot (Clontech) containing normalized loadings of poly(A)$^+$ RNA from 50 different human tissues was processed as above, except that the ExpressHyb™ hybridization buffer provided with the blot was used.

Immunoblotting

The Trx and His tags were cleaved from the fusion protein with thrombin and the resulting YSA1H protein with a residual N-terminal S-tag was used to raise a rabbit polyclonal antiserum by standard procedures. Cell lysates were separated by SDS/PAGE [13% (w/v) gel]. The gels were then equilibrated immediately in transfer buffer [10 mM Caps/NaOH (pH 11.0)/10% (v/v)

methanol] for at least 10 min before electrophoretic transfer of the separated proteins to a nitrocellulose membrane at 150 mA and 4 °C for 2 h. The membrane was blocked overnight at 4 °C with 3% (w/v) fat-free powdered milk and 0.2% (v/v) Tween 20 in PBS and then probed with a 1:1000 dilution of whole anti-YSA1H antiserum followed by a 1:5000 dilution of horseradish peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad). After being washed, the membrane was developed with the peroxidase substrate diaminobenzidine tetrahydrochloride (Sigma).

Other methods

Protein concentrations were estimated by the Coomassie Blue binding method [20]. N-terminal sequencing and electrospray mass spectrometric analysis were performed as described previously [6,18].

RESULTS

Cloning, expression and purification of YSA1H

A BLAST 2.0 search of the GenBank EST database with the *S. cerevisiae* YSA1 sequence yielded a large number of overlapping homologous human clones that, when assembled, revealed a 973-base sequence with a 657-base open reading frame (ORF) potentially encoding a 219-residue, 24.3 kDa protein, including the N-terminal methionine residue (Figure 1). This ORF had 63% amino acid sequence identity with the 26.1 kDa YSA1 protein over a 69-residue region encompassing the MutT motif (see Figure 6). We have named this putative protein human YSA1 homologue, or YSA1H. That the predicted sequence represents the full-length protein is indicated by the presence of

```
1                                                   31
CGC CCA TCC TTT TAG CAC CGC GGG AGG CGC CGG TGT TTC GAG CCG TGG ACC GGA TCG GCT
                  ***

61                                                  91
GAC ACT GCT GCC TCC AGC TAG TTA TTT CGT CCT CTT CCG TTC TTC ACC CCT ACA CCT TGG
                     ***

121                                 151             163/1
AGG TGA ACT TCT CAC CTG AGG GCT GTA AAG ACT CGT TTG AAA ATG GAG AGC CAA GAA CCA
 ***                                                M   E   S   Q   E   P

181/7                                               211/17
ACG GAA TCT TCT CAG AAT GGC AAA CAG TAT ATC ATT TCA GAG GAG TTA ATT TCA GAA GGA
T   E   S   S   Q   N   G   K   Q   Y   I   I   S   E   E   L   I   S   E   G

241/27                                              271/37
AAA TGG GTC AAG CTT GAA AAA ACA ACG TAC ATG GAT CCT ACT GGT AAA ACT AGA ACT TGG
K   W   V   K   L   E   K   T   T   Y   M   D   P   T   G   K   T   R   T   W

301/47                                              331/57
GAA TCA GTG AAA CGT ACA ACC AGG AAA GAG CAG ACT GCG GAT GGT GTC GCG GTC ATC CCC
E   S   V   K   R   T   T   R   K   E   Q   T   A   D   G   V   A   V   I   P

361/67                                              391/77
GTG CTG CAG AGA ACA CTT CAC TAT GAG TGT ATC GTT CTG GTG AAA CAG TTC CGA CCA CCA
V   L   Q   R   T   L   H   Y   E   C   I   V   L   V   K   Q   F   R   P   P

421/87                                              451/97
ATG GGG GGC TAC TGC ATA GAG TTC CCT GCA GGT CTC ATA GAT GAT GGT GAA ACC CCA GAA
M   G   G   Y   C   I   E   F   P   A   G   L   I   D   D   G   E   T   P   E

481/107                                             511/117
GCA GCT GCT CTC CGG GAG CTT GAA GAA GAA ACT GGC TAC AAA GGG GAC ATT GCC GAA TGT
A   A   A   L   R   E   L   E   E   E   T   G   Y   K   G   D   I   A   E   C

541/127                                             571/137
TCT CCA GCG GTC TGT ATG GAC CCA GGC TTG TCA AAC TGT ACT ATA CAC ATC GTG ACA GTC
S   P   A   V   C   M   D   P   G   L   S   N   C   T   I   H   I   V   T   V

601/147                                             631/157
ACC ATT AAC GGA GAT GAT GCC GAA AAC GCA AGG CCG AAG CCA AAG CCA GGG GAT GGA GAG
T   I   N   G   D   D   A   E   N   A   R   P   K   P   K   P   G   D   G   E

661/167                                             691/177
TTT GTG GAA GTC ATT TCT TTA CCC AAG AAT GAC CTG CTG CAG AGA CTT GAT GCT CTG GTA
F   V   E   V   I   S   L   P   K   N   D   L   L   Q   R   L   D   A   L   V

721/187                                             751/197
GCT GAA GAA CAT CTC ACA GTG GAC GCC AGG GTC TAT TCC TAC GCT CTA GCG CTG AAA CAT
A   E   E   H   L   T   V   D   A   R   V   Y   S   Y   A   L   A   L   K   H

781/207                                             811/217
GCA AAT GCA AAG CCA TTT GAA GTG CCC TTC TTG AAA TTT TAA GCC CAA ATA TGA CAC TGG
A   N   A   K   P   F   E   V   P   F   L   K   F  ***             ***

841/281                                             871/291
CCA TTT TTG TAA ACG AGA CCA CCA GGC CTT CTT CAC TAA GAC TTT GTA TTC AAC TTA GTT
    ***                                             ***

901                                                 931
TAA TGT AGA TTT GCC ATT AGC TTT TTC GTA AAA TAA AAG CAC AGA ACA GAA AAA AAA AAA
***                                                 ***

961
AAA AAA AAA AAA A
```

Figure 1 Nucleotide sequence of the YSA1H cDNA and predicted amino acid sequence of the YSA1H protein

In-frame stop codons upstream and downstream of the YSA1H ORF are indicated with three asterisks. The presumed initiating ATG in its strong initiating context AAAATGG is underlined with dashes. The MutT motif is underlined; the invariant amino acids are shown in bold. The likely polyadenylation signal AAATAAAA is doubly underlined.

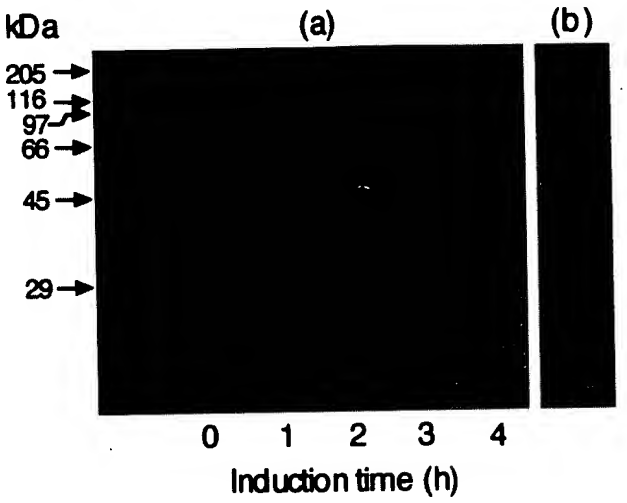


Figure 2 Overexpression and purification of Trx–YSA1H fusion protein

(**a**) *E. coli* BL-21(DE3) cells transformed with pET-YSA1H were induced with 1 mM IPTG for up to 4 h. Aliquots were taken at hourly intervals, boiled in sample buffer, analysed by SDS/PAGE [13% (w/v) gel] and stained with Coomassie Blue. (**b**) YSA1H protein after purification on Ni^{2+}/NTA-agarose as described in the Experimental section. The protein standards were myosin (205 kDa), β-galactosidase (116 kDa), phosphorylase *b* (97 kDa), BSA (66 kDa), ovalbumin (45 kDa) and carbonic anhydrase (29 kDa).

multiple stop codons both upstream and downstream of the ORF and by the highly favourable sequence context of the proposed initiator codon, AAAATGG (Figure 1) [21]. Human sequence-tagged sites corresponding to these ESTs have been mapped to chromosome 10p13–p14 between the markers D10S189 and D10S191 by the International Radiation Hybrid Mapping Consortium [22]. The UniGene entry number is Hs.11817.

A YSA1H cDNA was amplified by PCR from I.M.A.G.E. clone 310860, which was predicted to contain the full-length cDNA. For expression, the PCR fragment was inserted into the pET32b(+) vector and the recombinant plasmid (pET-YSA1H) was then used to transform *E. coli* BL-21(DE3) cells to generate a His-tagged thioredoxin fusion protein of predicted mass 41.6 kDa. After induction with IPTG, cell samples were analysed by SDS/PAGE, which showed that the soluble fraction contained a major IPTG-inducible protein band migrating with an apparent molecular mass of 47 kDa, which was therefore presumed to be the desired product (Figure 2a). The expressed Trx–YSA1H fusion protein was then purified in a single step to apparent homogeneity on a Ni^{2+}/NTA-agarose column (Figure 2b).

Substrate specificity

All MutT motif proteins studied so far have nucleotide pyrophosphatase activity that hydrolyses compounds containing an NDP-linked to another moiety. Therefore a range of nucleotides was assayed to determine the substrate specificity of Trx–YSA1H. In the presence of 5 mM Mg^{2+} ions, Trx–YSA1H was most active against ADP-sugars and some compounds of general structure ADP-sugar-X, e.g. Ap_2A and NADPH (Table 1). Lower activity was obtained with other NDP-sugars and CDP-alcohols, whereas no activity was obtained with (deoxy)nucleoside 5′-triphosphates, nucleoside 5′-diphosphates or nucleoside 5′-monophosphates or diadenosine polyphosphates, even when the enzyme concentration was increased

Table 1 Substrate specificities of YSA1H, human erythrocyte ADP-ribose pyrophosphatase, rat liver ADP-ribase II and calf liver ADP-sugar hydrolase

Results are expressed as the percentage activity obtained with ADP-ribose under the assay conditions used. Values in columns three to six are published (references in square brackets) and are based on different assay conditions from those used for YSA1H. Values for YSA1H are derived from assays containing 5 mM Mg^{2+} ions and are the averages of triplicate determinations. Abbreviation: n.t., not tested.

	Activity (% of value with ADP-ribose)				
		Human ADP-ribose pyrophosphatase		Rat liver ADP-ribase II [16]	Calf liver ADP-sugar hydrolase [17]
Substrate	YSA1H	Erythrocyte [15]	Erythrocyte [12]		
ADP-ribose	100	100	100	100	100
ADP-mannose	103	70	n.t.	n.t.	163
ADP-glucose	56	21	0	25	170
cADP-ribose	0	n.t.	0	n.t.	n.t.
Ap_2A	20	n.t.	n.t.	31	n.t.
NADPH	12	0	16	n.t.	n.t.
NADH	7	0	21	30	n.t.
NAD^+	7	0	42	< 2	5
Deamino-NAD^+	7	n.t.	n.t.	n.t.	n.t.
$NADP^+$	5	0	2	n.t.	n.t.
IDP-ribose	9	59	n.t.	31	n.t.
GDP-glucose	7	0	n.t.	n.t.	0
GDP-mannose	5	40	n.t.	< 2	n.t.
GDP-α-fucose	2	n.t.	n.t.	n.t.	n.t.
UDP-glucose	2	5	n.t.	< 2	0
UDP-galactose	7	0	n.t.	n.t.	n.t.
UDP-*N*-acetylglucosamine	< 1	n.t.	n.t.	n.t.	n.t.
CDP-glucose	3	5	n.t.	< 2	0
CDP-glycerol	3	n.t.	n.t.	2	n.t.
CDP-choline	< 1	n.t.	n.t.	< 2	n.t.
CDP-ethanolamine	< 1	n.t.	n.t.	< 2	n.t.
CoA	3	n.t.	n.t.	n.t.	n.t.
FAD	3	n.t.	n.t.	32	5
Ap_nA ($n = 3$–6)	0	0	n.t.	n.t.	n.t.
(d)NTP, (d)NDP, (d)NMP	0	0	n.t.	0	n.t.

20-fold. No activity was observed with thymidine-5′-monophospho-*p*-nitrophenyl ester as substrate; therefore YSA1H is devoid of alkaline phosphodiesterase 1 activity (EC 3.1.4.1). Comparative data are also presented in Table 1 for the human erythrocyte ADP-ribose pyrophosphatase, rat liver ADP-ribase II and calf liver ADP-sugar hydrolase. When assayed with 5 mM Mn^{2+} in place of Mg^{2+}, rat liver ADP-ribase II shows greatly increased activity with UDP-sugars and CDP-sugars and CDP-alcohols [16]. YSA1H activity towards CDP-glycerol was increased 4-fold when Mn^{2+} replaced Mg^{2+}, in comparison with the 50-fold increase observed with ADP-ribase II, whereas the activity of YSA1H towards ADP-sugars was decreased by 50–70% with Mn^{2+}. The UV-absorbing nucleotide reaction products generated from NDP-sugars were identified by HPLC and in each case were the corresponding NMP (Figure 3). The other product was assumed to be the sugar 5-phosphate.

Kinetic parameters and reaction requirements

Kinetic parameters were determined with ADP-ribose as substrate. K_m and V_{max} values were 60 μM and 8 μmol/min per mg respectively. The latter figure corresponds to a k_{cat} of 5.5 s^{-1}. With ADP-ribose as substrate, the enzyme displayed optimal activity at alkaline pH (7.4–9.0) with 2.5–5 mM Mg^{2+} or 100–250 μM Mn^{2+} ions (results not shown). In common with other MutT motif proteins, fluoride was inhibitory, with an IC_{50} of approx. 20 μM.

Tissue specificity of YSA1H mRNA expression

Northern blot and dot-blot analyses were performed to determine respectively the size and tissue distribution of YSA1H mRNA transcripts. A transcript of approx. 1.4 kb was detected in all human tissues present on the Northern blot (Figure 4a). A larger but less prominent transcript of approx. 5 kb was also detected, which might represent residual pre-mRNA as well as a smaller 1.1 kb species that was most prominent in placenta. Expression seemed stronger in lung, small intestine and stomach than in muscle, testis or placenta, but the blot was not normalized for differential rates of expression. A better comparison of tissue expression was obtained by dot-blot analysis with a blot that had been normalized with respect to eight different housekeeping genes. This showed that YSA1H was expressed in all 50 different human tissues present on the blot (Figure 4b). Although the amount of mRNA does not necessarily reflect expression at the protein level, such widespread expression is consistent with a role as a housecleaning enzyme required in all tissues. The level of mRNA expression was, however, higher in liver, kidney, pituitary, placenta and thymus than in other adult tissues and lowest in bone marrow, lymph nodes and certain areas of the brain.

Size and presence of YSA1H protein in human cell extracts

On gel-filtration analysis, the 41.6 kDa Trx–YSA1H fusion protein had a molecular mass of 85 kDa, suggesting that it

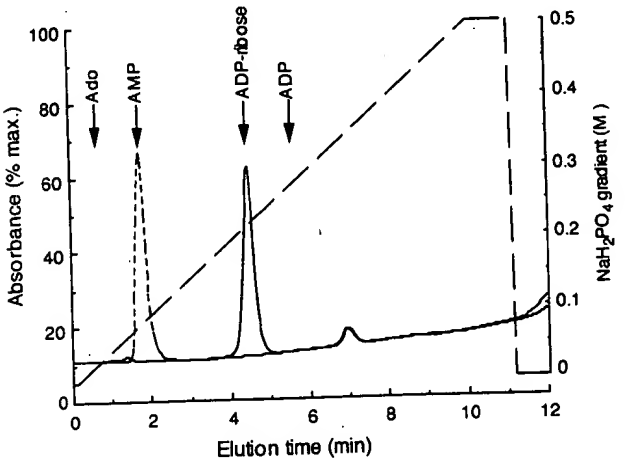


Figure 3 Identification of AMP reaction product from hydrolysis of ADP-ribose

Assays containing 0.3 mM ADP-ribose were incubated with or without 3 μg of Trx–YSA1H fusion protein for 30 min, then applied to a 1 ml Resource-Q anion-exchange column as described in the Experimental section. The positions of standards are indicated. The lines are as follows: solid line, without enzyme; short broken line, with enzyme; long broken line, gradient.

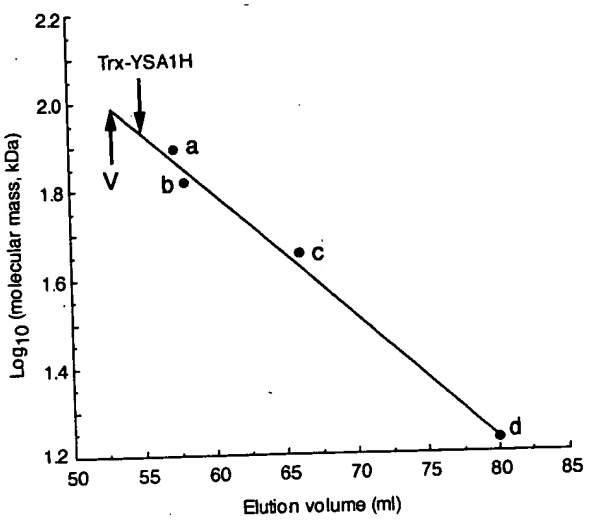


Figure 5 Gel-filtration analysis of purified recombinant Trx–YSA1H fusion protein

A sample of homogeneous Trx–YSA1H fusion protein was applied to a HiLoad 16/60 Superdex 75 column in 50 mM Tris/HCl (pH 7.5)/0.1 M NaCl and eluted at 1 ml/min in the same buffer. The column was calibrated with the following standards: a, conalbumin (77 kDa); b, BSA (66 kDa); c, ovalbumin (45 kDa); d, myoglobin (17 kDa). The void volume (V) and the elution position of Trx–YSA1H detected by its UV absorbance are indicated by arrows.

behaved as a dimer in solution (Figure 5). To show that the YSA1H protein was actually expressed in human cells, a rabbit polyclonal antibody was raised against recombinant YSA1H from which most of the vector fusion sequences had been removed by treatment with thrombin. Figure 6(a) shows an immunoblot analysis of the complete Trx–YSA1H fusion protein (lane 1) and after cleavage with thrombin, which yielded a product of theoretical molecular mass 27.5 kDa (lane 2), and enterokinase, which produced a near-native protein with a single alanine residue N-terminal to the initiating methionine residue (lane 3). Although the predicted molecular mass of the recombinant alanyl-YSA1H was 24.4 kDa, it ran on SDS/PAGE with

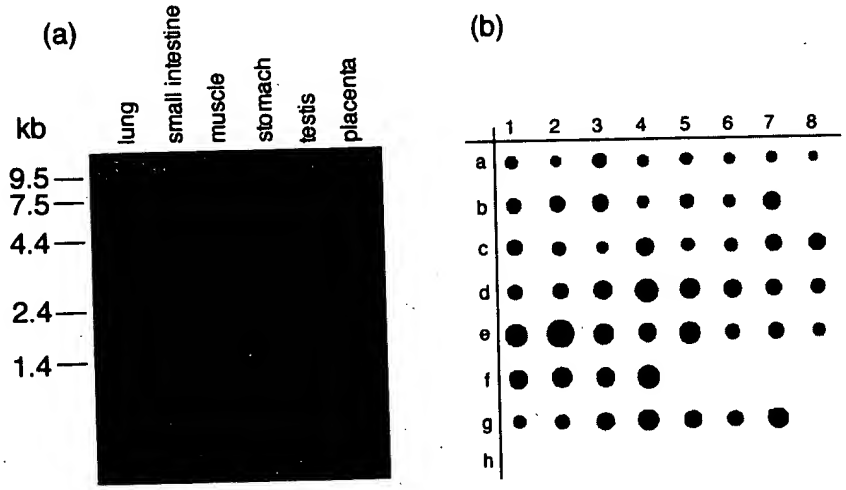


Figure 4 Expression of YSA1H mRNA in different human tissues

(**a**) Northern blot of poly(A)$^+$ RNA from five human tissues (2 μg per lane) probed with ^{32}P-labelled YSA1H anti-sense RNA transcribed *in vitro* from I. M. A. G.E. clone 310860 with T3 RNA polymerase. (**b**) RNA dot-blot containing normalized loadings of 89–514 ng of each poly(A)$^+$ RNA per dot from 50 different human tissues and 100 ng each of six control RNA and DNA species (500 ng for dot h8). The blot was probed as in (**a**). Dot identification: a1, whole brain; a2, amygdala; a3, caudate nucleus; a4, cerebellum; a5, cerebral cortex; a6, frontal lobe; a7, hippocampus; a8, medulla oblongata; b1, occipital lobe; b2, putamen; b3, substantia nigra; b4, temporal lobe; b5, thalamus; b6, nucleus accumbeus; b7, spinal cord; c1, heart; c2, aorta; c3, skeletal muscle; c4, colon; c5, bladder; c6, uterus; c7, prostate; c8, stomach; d1, testis; d2, ovary; d3, pancreas; d4, pituitary gland; d5, adrenal gland; d6, thyroid gland; d7, salivary gland; d8, mammary gland; e1, kidney; e2, liver; e3, small intestine; e4, spleen; e5, thymus; e6, peripheral leucocytes; e7, lymph node; e8, bone marrow; f1, appendix; f2, lung; f3, trachea; f4, placenta; g1, foetal brain; g2, foetal heart; g3, foetal kidney; g4, foetal liver; g5, foetal spleen; g6, foetal thymus; g7, foetal lung; h1. yeast total RNA; h2, yeast tRNA; h3, *E. coli* rRNA; h4, *E. coli* DNA; h5, poly(rA); h6, human C_0t 1 DNA; h7, human DNA; h8, human DNA.

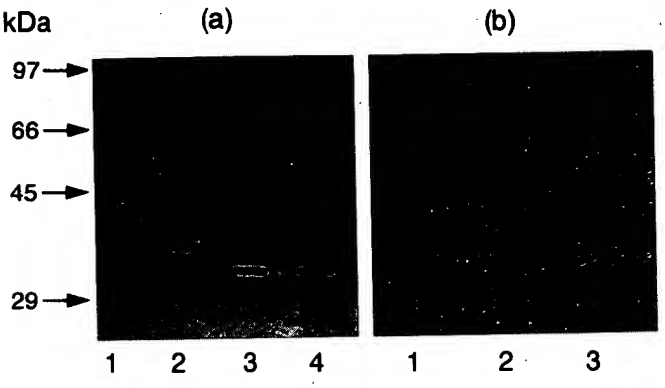


Figure 6 Immunoblot analysis of YSA1H protein expression in human cells

DAB-stained immunoblots were probed with anti-YSA1H as described in the Experimental section. (**a**) Lane 1, Trx–YSA1H fusion protein; lane 2, Trx–YSA1H cleaved with thrombin; lane 3, Trx–YSA1H cleaved with enterokinase; lane 4, KB cell extract. (**b**) Lane 1, U937 cell extract; lane 2, KB cell extract; lane 3, human erythrocyte extract.

an apparent molecular mass of 33 kDa. The N-terminal sequence of the enterokinase-cleaved recombinant protein was confirmed as AMESQEPT (including the vector alanine residue) by Edman degradation; electrospray mass spectrometric analysis of the same material confirmed its molecular mass as 24.4 kDa (results not shown). It therefore seems that recombinant YSA1H migrates anomalously on SDS/PAGE. This phenomenon of slower migration has been observed with other MutT motif proteins, including human Ap_4A hydrolase [6] and *S. cerevisiae* Ap_6A hydrolase [18] and might result from incomplete denaturation and SDS binding during sample preparation owing to the stability imparted by the mixed β-sheet core structure that might be a feature of this enzyme family [23].

The same species of apparent molecular mass 33 kDa was observed after examination of whole cell lysates prepared from human KB nasopharyngeal carcinoma cells (Figure 6a, lane 4, and Figure 6b, lane 2), human U937 promonocytic cells (Figure 6b, lane 1) and human erythrocytes (Figure 6b, lane 3). In addition, two extra bands of approx. 66 and 64 kDa were observed in the cell extracts. These might have been dimers resulting from incomplete sample denaturation; alternatively, they might have been unrelated proteins. We have previously identified two abundant proteins that cross-react strongly with an affinity-purified polyclonal antibody raised against another recombinant MutT motif protein, human Ap_4A hydrolase, as hsp60 (63 kDa) and glutamate dehydrogenase (61 kDa) (J. Kyte, J. L. Cartwright and A. G. McLennan, unpublished work). The purified human erythrocyte ADP-ribose pyrophosphatase has been reported to be a dimer of 34 kDa subunits [15]. If YSA1H is the same enzyme, our results would suggest that it is in fact a dimer of 24.3 kDa subunits.

Sequence comparisons

Compilation of mouse EST data revealed a 218-residue 24 kDa mouse homologue, mYSA1H, with overall 83% amino acid identity and 95% similarity to the human sequence. In addition to the *S. cerevisiae* YSA1 gene, closely related sequences are also present in rat and *Caenorhabditis elegans* (Figure 7). Significant matches to sequences from the prokaryotes *Treponema pallidum*, *Bacillus subtilis* (YQKG), *Synechocystis* sp. and to the *E. coli* YRFE gene (*orf186*) were also obtained (Figure 7). The *B. subtilis* YQKG gene has been reported to encode an ADP-ribose pyrophosphatase [10]; the *E. coli orf*186 gene product has a broad substrate specificity including ADP-ribose, NADH and Ap_3A [10]. The highly specific ADP-ribose pyrophosphatase from the archaeon *Methanococcus jannaschii* [13] and the *E. coli orf*209 gene product (YQIE, *rdsA*), which has also been reported to be an ADP-ribose pyrophosphatase [10], show only limited

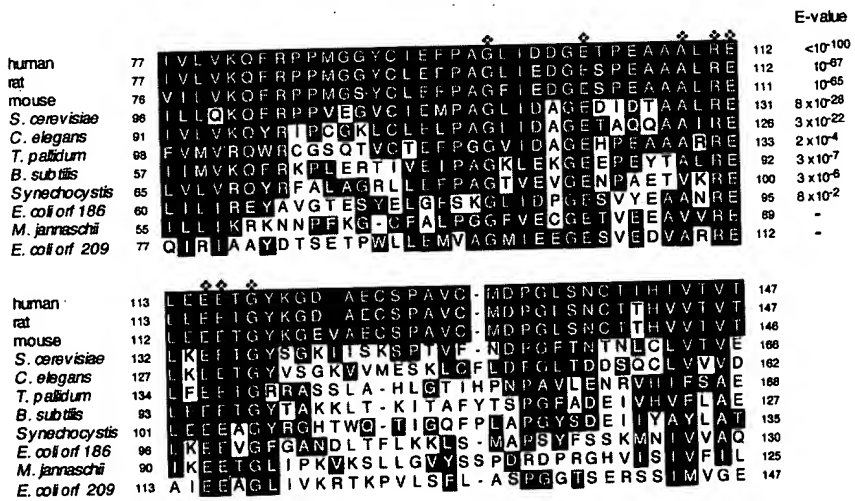


Figure 7 Multiple sequence alignment of YSA1H and related sequences

A 69-residue section of the YSA1H sequence encompassing the MutT motif was aligned by using the CLUSTAL W program with sequences from rat (AA858886), mouse (AI508660), *Saccharomyces cerevisiae* YSA1 (Q01976), *Caenorhabditis elegans* (AF067946), *Treponema pallidum* (O83713), *Bacillus subtilis* YQKG (P54570) and *Synechocystis* sp. (P72646), and the *Escherichia coli orf186* (P45799), *Methanococcus jannaschii* ADP-ribose pyrophosphatase (Q58549) and *E. coli orf209* (P36651) sequences; the numbers in parentheses are the accession numbers for these sequences. Amino acid identities with the human sequence are shaded black and similarities are shaded grey. 'Invariant' amino acids of the MutT motif are shown by ❖. The numbers to the right and left of the sequences are the amino acid positions in each complete sequence. Numbers in the right-hand column are the *E* values for these sequences from a BLAST 2.0 search of the GenBank EST and non-redundant databases, with the complete YSA1H as query sequence. The *E* value is a significance threshold representing the number of hits expected by chance when searching the database with the query sequence.

similarity to YSA1H overall but do show similarities in the region surrounding the MutT motif, in addition to the invariant amino acid residues of the motif itself, which might be involved in determining substrate specificity.

DISCUSSION

One mitochondrial and three cytosolic ADP-ribose pyrophosphatases have previously been partly purified from rat liver. The cytosolic ADP-ribase I and the mitochondrial enzyme are very similar in terms of low (micromolar) K_m for ADP-ribose and a stringent substrate specificity limited to ADP-ribose and IDP-ribose [16,24]. A similar activity has also been reported in embryonic cysts of the brine shrimp *Artemia franciscana* [25]. The cytosolic Mn^{2+}-dependent ADP-ribase has no activity at all with Mg^{2+} ions [16]. Therefore YSA1H does not seem to correspond to any of these enzymes. Its marked preference for ADP-sugar substrates indicates a closer similarity to the ADP-sugar pyrophosphatase activity detected in various mammalian liver extracts, including calf and rabbit, although the range of substrates analysed with this enzyme is rather limited [17]. There are also similarities to rat liver ADP-ribase II and the human erythrocyte ADP-ribose pyrophosphatase. These enzymes have a relatively high K_m for ADP-ribose (50–100 μM for rat, 170 μM for human erythrocyte and 60 μM for YSA1H) [15,16], whereas the native molecular mass of 47 kDa reported for rat ADP-ribase II [16] corresponds well to a dimer of 24.3 kDa subunits, the known subunit molecular mass of YSA1H. Apparent differences in substrate specificity include: (1) the high activity of the human erythrocyte enzyme with IDP-ribose and GDP-mannose [15]; (2) the lack of activity of the human erythrocyte enzyme with ADP-glucose [12] and (3) the high activity of the other enzymes with nicotinamide adenine dinucleotides. In the last case, significant contamination of commercial nucleotides with ADP-ribose was a likely cause of variation in results [16]. Because significant differences are apparent in the reported data for the same enzyme for human erythrocytes [12,15] and because rat ADP-ribase II exhibits a much broader specificity for NDP-sugars and nucleoside 5′-diphosphoalcohols in the presence of Mn^{2+} ions [16], differences in assay procedures and conditions might have a large role in the observed differences. Taken together, these results suggest that YSA1H, the human erythrocyte ADP-ribose pyrophosphatase, rat liver ADP-ribase II and mammalian liver ADP-sugar pyrophosphatase might be the same enzyme. However, this conclusion requires confirmation.

Fluoride is a potent inhibitor of several nudix hydrolases involved in diadenosine polyphosphate metabolism [7,18,26]. Because it also inhibits YSA1H with an IC_{50} of 20 μM, such inhibition might be a more general feature of this protein family. Because the specific rat liver and *Artemia* ADP-ribase I enzymes are similarly inhibited by fluoride [25], it is possible that they too are nudix hydrolases related to the specific *M. jannaschii* enzyme. The very stringent substrate specificity of these last three enzymes suggests that activities of this type might form the first line of defence against protein glycation by free ADP-ribose. However, ADP-sugars other than ADP-ribose are not significant metabolites in mammalian cells. ADP-glucose is an important precursor for bacterial glycogen and plant starch synthesis [27], whereas ADP-mannose has no known physiological function, although the commercially available synthetic compound can replace ADP-heptoses in bacterial outer-membrane lipopolysaccharide synthesis *in vitro* [28]. ADP-ribose might therefore also be the most important substrate for human YSA1H *in vivo*. Of the other potential substrates, Ap_2A is the most recent of the diadenosine polyphosphates to be identified *in vivo*. It is present in specific granules in human myocardium and can attain an intragranular concentration of 9 mM [29]. When released extracellularly, it acts as a coronary vasodilator. Within the granules, however, it might not be accessible to YSA1H and so might not be a significant substrate *in vivo*.

In conclusion, this study describes the biochemical properties of a new human member, YSA1H, of the MutT motif family of nucleotide pyrophosphatases and suggests a possible function. However, unequivocal confirmation of this function must await the results of genetic knock-out studies in yeast, *C. elegans* and mouse.

We thank M. C. Wilkinson for N-terminal sequencing and M. C. Prescott for mass spectrometric analysis. This work was supported by project grant 053038 from the Wellcome Trust and grant no. F25BJ from the Leverhulme Trust.

REFERENCES

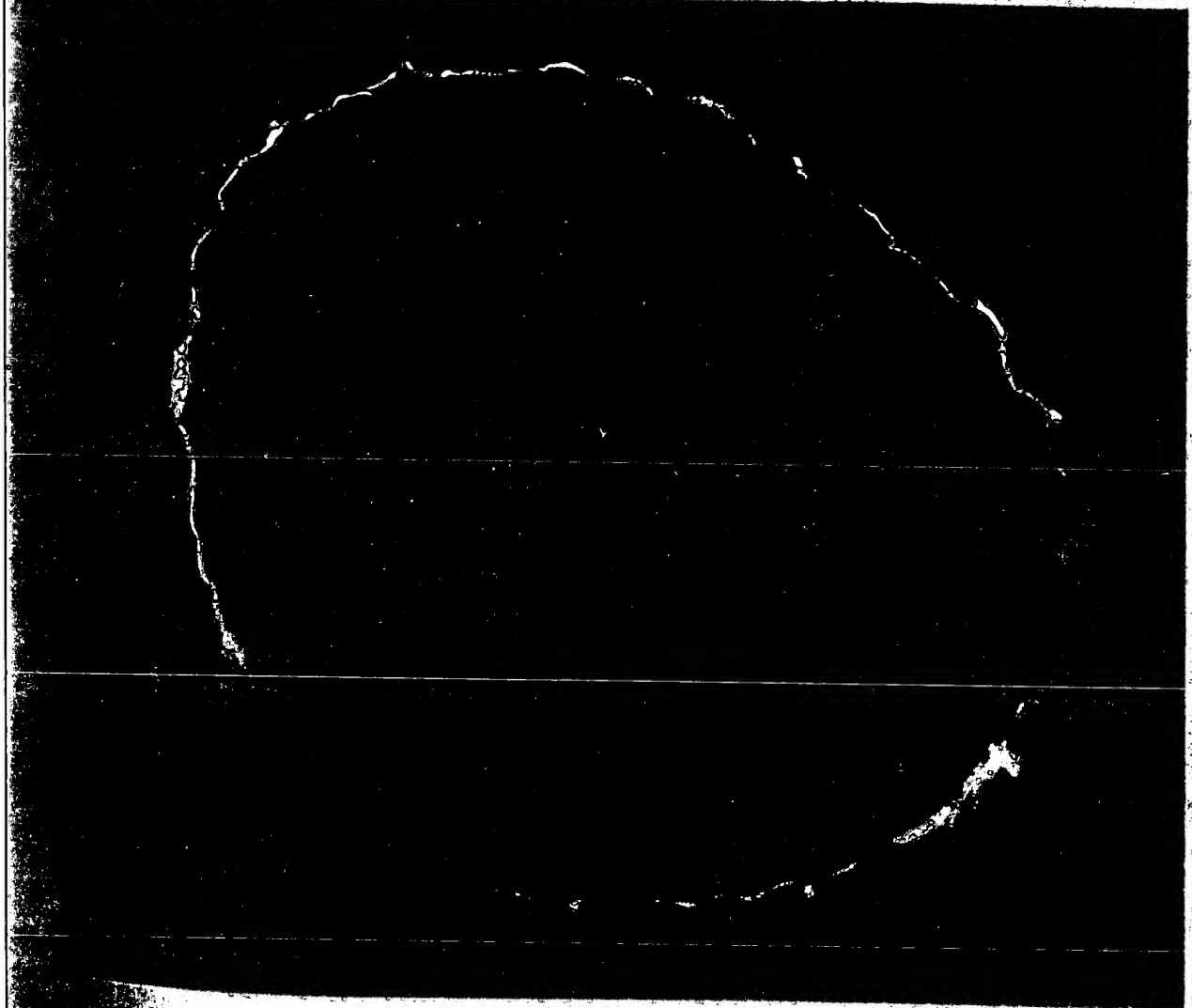
1 Maki, H. and Sekiguchi, M. (1992) Nature (London) **355**, 273–275
2 Koonin, E. V. (1993) Nucleic Acids Res. **21**, 4847
3 Bessman, M. J., Frick, D. N. and O'Handley, S. F. (1996) J. Biol. Chem. **271**, 25059–25062
4 McLennan, A. G. (1999) Int. J. Mol. Med. **4**, 79–89
5 Sakumi, K., Furuichi, M., Tzuzuki, T., Kakuma, T., Kawabata, S.-I., Maki, H. and Sekiguchi, M. (1993) J. Biol. Chem. **268**, 23524–23530
6 Thorne, N. M. H., Hankin, S., Wilkinson, M. C., Nunez, C., Barraclough, R. and McLennan, A. G. (1995) Biochem. J. **311**, 717–721
7 Safrany, S. T., Caffrey, J. J., Yang, X. N., Bembenek, M. E., Moyer, M. B., Burkhart, W. A. and Shears, S. B. (1998) EMBO J. **17**, 6599–6607
8 Safrany, S. T., Ingram, S. W., Cartwright, J. L., Falck, J. R., McLennan, A. G., Barnes, L. D. and Shears, S. B. (1999) J. Biol. Chem. **274**, 21735–21740
9 Li, A. W., Too, C. K. L. and Murphy, P. R. (1996) Biochem. Biophys. Res. Commun. **223**, 19–23
10 O'Handley, S. F., Frick, D. N., Dunn, C. A. and Bessman, M. J. (1998) J. Biol. Chem. **273**, 3192–3197
11 Moss, J. and Vaughan, M. (1990) ADP-ribosylating Toxins and G Proteins, American Society for Microbiology, Washington, DC
12 Zocchi, E., Guida, L., Franco, L., Silvestro, L., Guerinni, M., Benatti, U. and De Flora, A. (1993) Biochem. J. **295**, 121–130
13 Sheikh, S., O'Handley, S. F., Dunn, C. A. and Bessman, M. J. (1998) J. Biol. Chem. **273**, 20924–20928
14 Kwak, Y.-G., Park, S.-K., Kim, U.-H., Han, M.-K., Eun, J.-S., Cho, K.-P. and Chae, S.-W. (1996) Am. J. Physiol. **271**, C464–C468
15 Kim, J. S., Kim, W. Y., Rho, H. W., Park, J. W., Park, B.-H., Han, M. K., Kim, U. H. and Kim, H. R. (1998) Int. J. Biochem. Cell Biol. **30**, 629–638
16 Canales, J., Pinto, R. M., Costas, M. J., Hernández, M. T., Miró, A., Bernet, D., Fernández, A. and Cameselle, J. C. (1995) Biochim. Biophys. Acta **1246**, 167–177
17 Rodriguez, P., Bass, S. T. and Hansen, R. G. (1968) Biochim. Biophys. Acta **167**, 199–201
18 Cartwright, J. L. and McLennan, A. G. (1999) J. Biol. Chem. **274**, 8604–8610
19 Gasmi, L., Cartwright, J. L. and McLennan, A. G. (1998) Biochim. Biophys. Acta **1405**, 121–127
20 Peterson, G. L. (1982) Methods Enzymol. **91**, 95–119
21 Kozak, M. (1996) Mammal. Genome **7**, 563–574
22 Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tomé, P., Hui, L., Matise, T. C., McKusick, K. B., Beckmann, J. S. et al. (1998) Science **282**, 744–746
23 Abeygunawardana, C., Weber, D. J., Gittis, A. G., Frick, D. N., Lin, J., Miller, A. F., Bessman, M. J. and Mildvan, A. S. (1995) Biochemistry **34**, 14997–15005
24 Bernet, D., Pinto, R. M., Costas, M. J., Canales, J. and Cameselle, J. C. (1994) Biochem. J. **299**, 679–682
25 Fernández, A., Ribeiro, J. M., Costas, M. J., Pinto, R. M., Canales, J. and Cameselle, J. C. (1996) Biochim. Biophys. Acta **1290**, 121–127
26 Guranowski, A. (1990) FEBS Lett. **262**, 205–208
27 Smith-White, B. J. and Preiss, J. (1992) J. Mol. Evol. **34**, 449–464
28 Kadrmas, J. L., Brozek, K. A. and Raetz, C. R. H. (1996) J. Biol. Chem. **271**, 32119–32125
29 Luo, J., Jankowski, J., Knobloch, M., van der Giet, M., Gardanis, K., Russ, T., Vahlensieck, U., Neumann, J., Schmitz, W. et al. (1999) FASEB J. **13**, 695–705

Received 8 July 1999/20 August 1999; accepted 15 September 1999

:258 (print)
3-351x (electronic)
277(21) 18245-19242(2002)

Journal of Biological Chemistry

MAY 24, 2002 VOLUME 277 NUMBER 21



PUBLISHED BY THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY

Founded by Christian A Herter and Sustained in Part by the Christian A Herter Memorial Fund

...AL OF BIOLOGICAL CHEMISTRY
...by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 277, No. 21, Issue of May 24, pp. 18579–18585, 2002
Printed in U.S.A.

...at Mast Cell Protease 4 Is a β-Chymase with Unusually Stringent ...bstrate Recognition Profile*

Received for publication, October 29, 2001, and in revised form, March 8, 2002
Published, JBC Papers in Press, March 14, 2002, DOI 10.1074/jbc.M110356200

Ulrika Karlson‡, Gunnar Pejler§, Gunnar Fröman¶, and Lars Hellman‡||

From the ‡Department of Cell and Molecular Biology, Uppsala University, The Biomedical Center, SE-751 24 Uppsala, Sweden, the §Swedish University of Agricultural Sciences, Department of Veterinary Medical Chemistry, The Biomedical Center, SE-751 23 Uppsala, Sweden, and the ¶Department of Medical Biochemistry and Microbiology, Uppsala University, The Biomedical Center, SE-751 23 Uppsala, Sweden

...ated mast cells release a variety of potent inflam...tory mediators including histamine, cytokines, pro...glycans, and serine proteases. The serine proteases ...y to either the chymase (chymotrypsin-like sub... specificity) or tryptase (trypsin-like specificity) ...ily. In this report we have investigated the substrate ...ificity of a recently identified mast cell protease, rat ...t cell protease-4 (rMCP-4). Based on structural ho...logy, rMCP-4 is predicted to belong to the chymase ...ily, although rMCP-4 has previously not been char...terized at the protein level. rMCP-4 was expressed ...th an N-terminal His tag followed by an enterokinase ...e substituting for the native activation peptide. The ...terokinase-cleaved fusion protein was labeled by di...propyl fluorophosphate, demonstrating that it is an ...ive serine protease. Moreover, rMCP-4 hydrolyzed ...eO-Suc-Arg-Ala-Tyr-pNA, thus verifying that this pro...ase belongs to the chymase family. rMCP-4 bound to ...parin, and the enzymatic activity toward MeO-Suc-...g-Ala-Tyr-pNA was strongly enhanced in the presence ...heparin. Detailed analysis of the substrate specificity ...as performed using peptide phage display technique. ...ter six rounds of amplification a consensus sequence, ...-Val-Trp-Phe-Arg-Gly, was obtained. The corre...onding peptide was synthesized, and rMCP-4 was ...own to cleave only the Phe-Arg bond in this peptide. ...is demonstrates that rMCP-4 displays a striking pref...ence for bulky/aromatic amino acid residues in both ...P1 and P2 positions.

...st cells are potent inflammatory cells with well known ...ful effects, *e.g.* during allergic reactions. However, their ...ary beneficial function *in vivo* is most likely their involve...t in the expulsion of nematode parasites and in the defense ...st bacterial infections (1–3). Two main types of mast cells ...been identified in rodents, the connective tissue mast ...which mainly reside in connective tissue and the perito...m, and the mucosal mast cells (MMC),[1] which reside ...ly in the mucosa of the respiratory and intestinal tracts. Both types of rodent mast cells are activated by antigen cross-linking of surface-bound IgE and respond by rapidly releasing mediators stored in their granules. The inflammatory mediators include histamine, proteoglycans, prostaglandins, cytokines, and neutral proteases (4). Mast cells are a particularly rich source of serine proteases belonging to the tryptase family, the chymase family, and the recently identified mMCP-8 family. Proteases belonging to the mMCP-8 family have as yet unknown substrate specificity. The tryptase family of proteases cleaves their substrate at the C-terminal side of basic amino acids (Arg and Lys), whereas the chymase family has chymotrypsin-like substrate specificity, *i.e.* they cleave peptides after aromatic amino acids (Phe, Tyr, and Trp). Based on phylogenetic relationships, the chymase family can be further subdivided into two groups, the α- and the β-chymases (5). In nonrodent mammals only a single α-chymase can be found, whereas in rodents several β-chymases are expressed in addition to the α-chymase. At least 18 mammalian α- and β-chymases have been identified and cloned (6). However, only a few of them have been characterized in detail with regard to structure and cleavage specificity.

Five of the ten rat mast cell proteases (rMCP-1, rMCP-2, rMCP-3, rMCP-4, and rMCP-5) are considered to display chymase activity (7). rMCP-5 (initially designated as rMCP-3) is the rat α-chymase and the homologue to the single human chymase found so far. The rat mast cell β-chymases rMCP-1 and rMCP-2, expressed by connective tissue mast cells and MMC, respectively, have been extensively investigated with regard to tissue distribution, charge, heparin binding, substrate specificity, inhibitor susceptibility, and structure (7–10). Recent data suggest that MMC mediators, in particular rMCP-2, directly induce physiological changes in the gastrointestinal tract, such as an increase in epithelial permeability (11). This results in increased leukocyte migration and microbial product influx that potentially enhances the inflammatory response.

rMCP-3 and rMCP-4 (GenBank™ accession numbers U67888 and U67907, respectively) were initially cloned from the rat mucosal mast cell line RBL-1 (7) and group phylogenetically with the β-chymases. Both rMCP-3 and rMCP-4 have been suggested to be MMC proteases, with detectable mRNA levels primarily found during parasitic infections (12). However, neither rMCP-3 nor rMCP-4 has been characterized at the protein level. Chymases utilize extended substrate-binding sites to discriminate between substrates. Additional understanding of the biological role of a protease may be obtained by a detailed analysis of its substrate specificity. In the present

...work was supported by grants from the Swedish Natural ... Research Council, the Swedish Medical Research Council, and ...ltelsen. The costs of publication of this article were defrayed in ...the payment of page charges. This article must therefore be ...marked "*advertisement*" in accordance with 18 U.S.C. Section ...ely to indicate this fact.

...om correspondence should be addressed: Department of Cell ...olecular Biology, Uppsala University, The Biomedical Center, ...SE-751 24 Uppsala, Sweden. Tel.: 46-18-471-4532; Fax: 46-18-...34; E-mail: lars.hellman@icm.uu.se.

...abbreviations used are: MMC, mucosal mast cell; DFP, diiso...fluorophosphate; EK, enterokinase; mMCP, mouse mast cell pro...tease; rMCP, rat mast cell protease; PBS, phosphate-buffered saline; Ni-NTA, nickel-nitrilotriacetic acid; HPLC, high pressure liquid chromatography.

study, rMCP-4 has been characterized with regard to activity, heparin binding, and extended substrate specificity.

EXPERIMENTAL PROCEDURES

Subcloning, Expression, and Purification of Recombinant rMCP-4—The rMCP-4 expression construct was obtained by PCR amplification using full-length cDNA (7) as template. For purification purposes the 5′-primer contained a sequence encoding six histidine residues. The 5′-primer also introduced an enterokinase susceptible peptide (Asp-Asp-Asp-Asp-Lys) in replacement of the natural activation peptide. The enterokinase cleavage site enables subsequent removal of the histidine residues and thereby enables activation of the protease. The PCR product was inserted into the pCEP-Pu2 vector (13), and the nucleotide sequence of the final product was confirmed using the Thermo Sequenase-radiolabeled terminator cycle sequencing kit (Amersham Biosciences) and vector specific primers. Human embryonic kidney cells, 293-EBNA, were transfected with the pCEP-Pu2/rMCP-4 construct as previously described (13). Selection was initiated by the addition of 5 μg/ml puromycin to the cell culture medium (Dulbecco's modified Eagle's medium supplied with 10% fetal calf serum, 2 mM L-glutamine, and 50 μg/ml gentamicin). After 1 week of selection, the level of puromycin was decreased to 0.5 μg/ml. The conditioned medium was collected and centrifuged to remove cell debris, followed by the addition of 0.5 ml of Ni-NTA-agarose beads/liter (Qiagen). After 6 h of incubation at 4 °C on a shaker, the beads were pelleted by centrifugation and transferred to 10-ml PolyPrep Chromatography columns (Bio-Rad). The beads were washed with PBS (pH 7.2) containing 1 M NaCl and 0.1% Tween 20, and the column was eluted with 100 mM imidazole in PBS.

Protein purity and concentration was estimated on SDS-PAGE. The samples were mixed with sample buffer and 5% β-mercaptoethanol and run on a 12% gel. To visualize protein bands, the gel was stained with Coomassie Brilliant Blue.

Activation and Active Site Titration of rMCP-4—Recombinant rMCP-4 was digested for 5h with EKMax™ (Invitrogen) enterokinase, one unit/10 μg of recombinant protease. A 10-ml PolyPrep Chromatography column containing 0.2 ml of heparin-Sepharose (Amersham Biosciences) was equilibrated with PBS (pH 7.2). 10 μg of EK-cleaved rMCP-4 in 125 μl of PBS (pH 7.2) was applied to the column, followed by washing with PBS (0.15 M NaCl) and stepwise elution with increasing NaCl concentrations from 0.25 to 1.15 M. The flow-through and eluted fractions were assayed by the chromogenic substrate S-2586 and run on SDS-PAGE to estimate the protein content.

Initial experiments showed that rMCP-4 bound to heparin-Sepharose. To purify rMCP-4 from enterokinase and other impurities, EK-digested rMCP-4 samples were therefore purified by affinity chromatography on heparin-Sepharose (see above). Following affinity chromatography, the protein concentration was determined by measuring A_{280} with a calculated extinction coefficient of 20690 M^{-1} cm^{-1}. Enzymatic activity was measured toward the chromogenic substrate S-2586 (MeO-Suc-Arg-Ala-Tyr-pNA) (Chromogenix, Mölndal, Sweden). Standard measurements were performed in 96-well microtiter plates with a substrate concentration of 0.18 mM in 200 μl of PBS. S-2586 hydrolysis was monitored spectrophotometrically at 405 nm in a Multiscan MCC/340 spectrophotometer (Labsystem). The dependence on heparin for enzymatic activity was assayed after the addition of pig mucosal heparin (Sigma) at different enzyme-to-heparin mass ratios. As controls, the samples with heparin or enterokinase only were included. In addition, the activity of rMCP-4 that had not been EK-digested was measured. To evaluate the fraction of active rMCP-4, the protease (1 μg) was titrated with human α_1-antichymotrypsin (Calbiochem). The residual activity was measured with S-2586 after overnight incubation of rMCP-4 with the inhibitor at different inhibitor to chymase molar ratios.

Binding of the Active Site Inhibitor, Diisopropylfluorophosphate—Three samples, each containing ~2 μg of His_6-EK-rMCP-4 in PBS were prepared. 10 μg of pig mucosal heparin and 0.7 μl (0.7 unit) of EK Max enterokinase were added to one of the samples. EK (0.7 μl) or heparin (10 μg) was added to the remaining two samples, respectively. Following 5 h of incubation at 37 °C, the aliquots were mixed with 5 μCi of [^{3}H] DFP (PerkinElmer Life Sciences) and incubated for 16h at 37 °C. After incubation, the samples were assayed by SDS-PAGE. The gel was stained with Coomassie Brilliant Blue, incubated for 1 h in 1 M sodium salicylate, and dried in a vacuum dryer. To analyze the amount of labeled DFP covalently bound to the protease active site, an x-ray film was exposed to the gel for 2 weeks.

Determination of Cleavage Recognition Profile by a Phage Display Approach—A library of substrate peptides was generated by synthesizing degenerate oligonucleotides with the nucleotide sequence AATTC[illegible] ACTCCAGGCGGC-$(NNK)_9$-CATCACCATCACCATCACTAA[illegible] (N) sents any nucleotide, and K represents T or G) encoding a [illegible] nonamer followed by a histidine tag. The oligonucleotides were in[illegible] into the coding region of the T7 phage capsid protein in the T7S[illegible] vector arms (T7Select system manual; Novagen) by the method [illegible] to Cwirla *et al.* (14). The vector constructs were packaged into [illegible] particles *in vitro*. The total number of T7 clones after packaging was [illegible] 10^7 plaque forming units. The phage library was amplified to a titer [illegible] 10^{10} plaque forming units/ml in the BLT 5615 bacterial strain. Am[illegible] cation is necessary for the expression of His-tagged inserts and [illegible] display on the surface of the phage particles. The His-tagged [illegible] protein is only produced in very low numbers (0.1–1/phage) because [illegible] upstream partial deletion of the capsid protein promoter region [illegible] T7Select 1–1 vector. Hence, a His-tagged capsid protein is present [illegible] least 10% of the assembled phage particles. The native capsid prot[illegible] expressed by the host bacterial strain, BLT 5615, upon isopropyl-1-[illegible] β-D-galactopyranoside induction.

An aliquot of the amplified phages (~10^9 plaque forming units) [illegible] allowed to bind to 100 μl of Ni-NTA-agarose beads for 2 h while [illegible] gently. Unbound phages were removed by washing 10 times with 1 [illegible] of 1 M NaCl, 0.1% Tween 20 in PBS, pH 7.2, further washed twice [illegible] ml of PBS, finally resuspended in 100 μl of PBS. A control elution of [illegible] phages with 100 μl of 500 mM imidazole concluded that at least 1[illegible] phages were attached to the matrix after washing. rMCP-4 was [illegible] vated as previously described in the absence of heparin and then [illegible] jected to heparin-Sepharose purification. Following elution with [illegible] containing 0.5 M NaCl, heparin was added to a 1:5 mass ratio [illegible] protease to heparin. The selection was started by adding 0.5 [illegible] protease (200 nM) or buffer without protease as a control to the [illegible] with the resuspended beads. The protease was allowed to digest [illegible] ceptible phages at 37 °C overnight with gentle agitation. To recover [illegible] released phages, the Ni-NTA-agarose beads were pelleted by centrifu[illegible] gation, and the phages in the supernatant were removed in a [illegible] volume of 3 × 100 μl of PBS (pH 7.2). To ensure that all recovered [illegible] phages lacked the histidine tag, 15 μl of fresh Ni-NTA-agarose bead[illegible] were added to the phage suspension, and the mixture was agitated [illegible] 15 min followed by centrifugation to recover the supernatant. 10 μl [illegible] the supernatant was used to determine the amount of detached phages [illegible] in each round of selection. The remaining 290 μl of the supernatant was [illegible] added to a 10-ml culture (OD = ~0.6) of *Escherichia coli* (BLT5[illegible] The bacteria had been induced with 100 μl of 100 mM isopropyl-1-[illegible] β-D-galactopyranoside 30 min before phage addition to ensure [illegible] tion of the phage capsid protein. Approximately 2 h after phage [illegible] tion, the bacteria lysed, and the phage sublibrary was added to [illegible] agarose beads. After binding and washing the sublibrary, a new [illegible] of selection was started. Following six rounds of selection, 35 plaques were arbitrarily isolated from LB plates, and each dissolved in phage extraction buffer (100 mM NaCl, 20 mM Tris-HCl, pH 8.0, and 6 mM $MgSO_4$). To disrupt the phages, the dissolved material was mixed with 10 mM EDTA, pH 8.0, and heated at 65 °C for 10 min. The phage DNA was then amplified by PCR, using T7 SelectUP and T7SelectDOWN primers (T7Select Cloning kit, Novagen). After amplification the PCR fragments were purified using the QIAquick PCR purification kit (Qiagen). The purified PCR fragments were then sequenced using Thermo Sequenase-radiolabeled terminator cycle sequencing kit (Amersham Biosciences). [illegible] 2. In a control experiment, to verify the completeness and functionality of the library, 1.5 μl (1 unit) of thrombin (Sigma) was used to [illegible] library, because the substrate specificity for thrombin is known [illegible] The protocol for the selections with thrombin was the same as [illegible] rMCP-4, except that the incubation time was decreased to 2 h, and [illegible] four rounds of selection were performed.

Testing Phage Display Results in a Synthetic Peptide—To verify that the sequences from the phage display selections were indeed substrates for rMCP-4, and to confirm the cleavage site (P1–P1′ amino acids), a synthetic peptide was constructed from the aligned sequences. The peptide substrate, NH_2-LVWFRG-COOH (purchased from [illegible] Engström, Department of Medical Biochemistry and Microbiology, BMC, Uppsala, Sweden), was dissolved in Me_2SO. Different concentrations of the peptide (10% Me_2SO in PBS, pH 7.2) were incubated for [illegible] min at 37 °C with 16 ng of rMCP-4, with or without heparin. [illegible] samples from the incubation mixtures were applied to a Source 5RPC ST 4.6/150 HPLC column (Amersham Biosciences) pre-equilibrated in a 2:3 ratio (v/v) of 70% acetonitrile with 10 mM NaOH. The column was eluted with a linear gradient of acetonitrile ranging from 30 to 70% over 25 min. The effluent was analyzed by monitoring spectrophotometri[illegible]

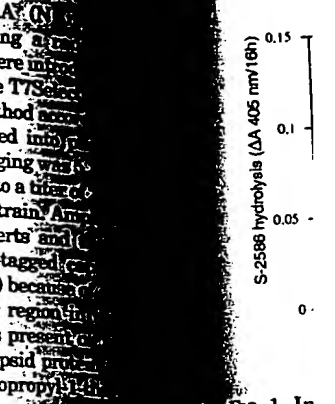


FIG. 1. In[illegible] raphy of [illegible] ked to a h[illegible] with increa[illegible] fractions w[illegible] stimulation [illegible] mixed with [illegible] zymatic [illegible]

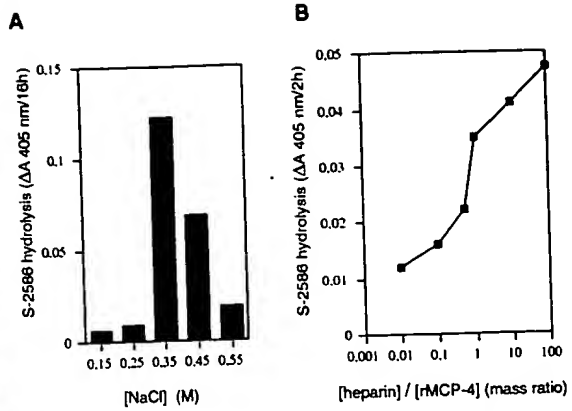


FIG. 1. **Interaction of rMCP-4 with heparin.** *A*, affinity chromatography of rMCP-4 on heparin-Sepharose. A sample (10 μg) was applied to a heparin-Sepharose column. The column was eluted stepwise with increasing NaCl concentrations. Enzymatic activity in the eluted fractions was measured with the chromogenic substrate S-2586. *B*, stimulation of rMCP-4 activity toward S-2586 by heparin. rMCP-4 was mixed with pig mucosal heparin at different mass (w/w) ratios. The enzymatic activity was measured with S-2586.

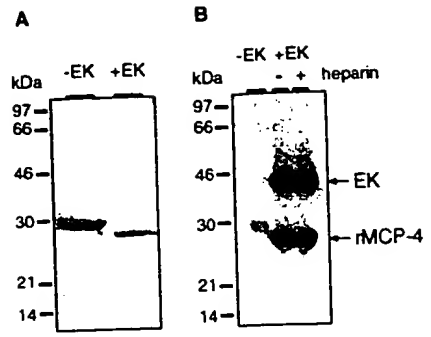


FIG. 2. **Expression of recombinant rMCP-4 by HEK 293-EBNA cells.** *A*, His-tagged recombinant protease was purified from cell culture medium by using Ni-NTA-agarose beads, as described under "Experimental Procedures." The purified protein sample was then cleaved with enterokinase to remove the purification tag. Subsequently, enterokinase-digested rMCP-4 was purified on heparin-Sepharose to remove the enterokinase. rMCP-4 digested with enterokinase (+EK) and one control sample without enterokinase digestion (−EK) were analyzed by SDS-PAGE (~2.5 μg of protein/lane) followed by staining with Coomassie Brilliant Blue. *B*, binding of DFP to recombinant rMCP-4. Approximately 2 μg of EK-cleaved rMCP-4 (+EK) with (+) or without (−) heparin and the same amount of uncleaved rMCP-4 (−EK) were incubated with [^{3}H]DFP as described under "Experimental Procedures." The samples were separated by SDS-PAGE (12% polyacrylamide), and [^{3}H]DFP-labeled bands were visualized by autoradiography.

cally at 412 nm. Fractions corresponding to peaks were collected and analyzed by mass spectroscopy.

RESULTS

Expression, Purification, and Analysis of Recombinant rMCP-4—The coding region for rMCP-4 was PCR-amplified using the corresponding full-length cDNA clone as template. The construct contained an N-terminal His_6 tag followed by an EK site that replaced the natural activation peptide. The resulting fragment of 711 bp was inserted in the pCEP-Pu2 vector in frame with the BM40 signal sequence, which is a signal sequence derived from the human protein SPARC (secreted protein, acidic and rich in cysteine), also referred to as osteonectin. The pCEP-Pu2 vector containing the protease construct was introduced into the 293-EBNA cell line by LipofectAMINE transfection. Recombinant rMCP-4 present in the culture medium was purified by Ni-NTA affinity chromatography and was subjected to enterokinase digestion. Many of the various isolated chymases show binding to heparin, mediated by positively charged amino acid residues located at the chymase surface (16–18). Experiments were therefore carried out to determine whether the recombinant rMCP-4 showed affinity for heparin. EK-digested rMCP-4 was loaded on columns of heparin-Sepharose, which were eluted stepwise with buffers containing increasing concentrations of NaCl. These experiments showed almost 100% binding of rMCP-4 to heparin-Sepharose at 0.15 M NaCl and elution of the protein at ~0.35 M NaCl (Fig. 1*A*). The heparin-Sepharose affinity was utilized for the purification of rMCP-4 after the EK step, allowing separation of EK and other impurities from the EK-digested mature rMCP-4 protein. [illegible] The molecular mass of rMCP-4 before EK digestion, as determined by the migration on SDS-PAGE, was estimated to be [illegible] kDa (Fig. 2*A*). After enterokinase digestion a molecular mass of 26 kDa was observed for the mature protease, which correlates well with the calculated theoretical value of 24,987 Da. Because the histidine tag and the enterokinase cleavage site add ~1.5 kDa in size to the recombinant protease, the molecular mass of the uncleaved fusion protein is close to the expected value.

Activity Analysis of Recombinant rMCP-4—Based on structural comparisons with other serine proteases, rMCP-4 is expected to display chymotrypsin-like substrate specificity, *i.e.* cleaving substrates at the C-terminal side of aromatic amino acid residues. To test whether this is the case, the ability of rMCP-4 to hydrolyze S-2586 (MeO-Suc-Arg-Ala-Tyr-pNA) was tested. S-2586 has previously been utilized as an efficient substrate for the main chymase found in rat connective tissue mast cells, rMCP-1 (19). rMCP-4 was indeed shown to hydrolyze S-2586. S-2586 hydrolysis was strongly enhanced by heparin in a dose-dependent manner and was barely detectable at suboptimal heparin concentrations (Fig. 1*B*). The specific activity toward S-2586 (at 1 mM concentration of S-2586) was ~0.6 pmol of hydrolyzed S-2586 $s^{-1}/\mu g$ rMCP-4 at a 1:1 protease to heparin (w/w) ratio. Attempts were made to determine K_m and k_{cat} values of rMCP-4 for S-2586. However, no leveling out of the curve corresponding to the substrate cleaving rates was reached at S-2586 concentrations up to 3.5 mM. Above this concentration S-2586 was not soluble, and determination of kinetic parameters was therefore not possible.

The serine protease inhibitor DFP, which covalently binds to the serine residue in the catalytic triad, can be used to detect active serine proteases. EK-digested rMCP-4, with or without heparin, was incubated with ^{3}H-labeled DFP. The samples were separated by SDS-PAGE, followed by autoradiography. Fig. 2*B* shows that EK-digested rMCP-4 binds DFP, both in the presence and absence of heparin, whereas its non-EK-digested counterpart does not. This shows that rMCP-4 is an active protease. Further, because no additional DFP-binding bands other than rMCP-4 and EK were visible (removal of enterokinase by heparin chromatography was not done in the DFP labeling experiment), the rMCP-4 preparation was free from other contaminating serine proteases. To confirm that no other chymotryptic proteases were present in the purified preparation, the activity toward S-2586 was tested in samples not activated with enterokinase. No activity toward S-2586 was detected in such samples (data not shown).

Active Site Titration of rMCP-4—EK-digested rMCP-4 was active site-titrated with α_1-antichymotrypsin, a protease inhibitor belonging to the serpin family. Close to 100% inhibition of rMCP-4 was obtained at a molar ratio of inhibitor to protease of ~1:1, indicating that essentially all of the recombinant rMCP-4 molecules were correctly folded (Fig. 3).

Determination of Substrate Recognition Profile by Substrate Phage Selection—A phage display analysis was performed to determine the substrate specificity of rMCP-4 in more detail.

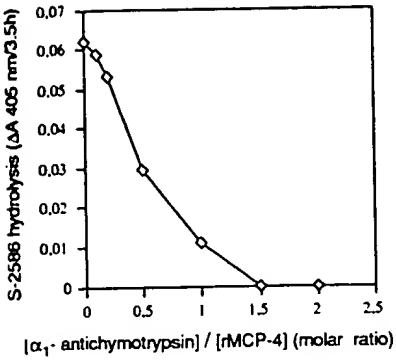


FIG. 3. **Titration of rMCP-4 with α_1-antichymotrypsin.** α_1-Antichymotrypain was added to samples of rMCP-4 (1 μg) at the various molar ratios indicated. After 16 h of incubation, residual activities were measured with the chromogenic substrate S-2586.

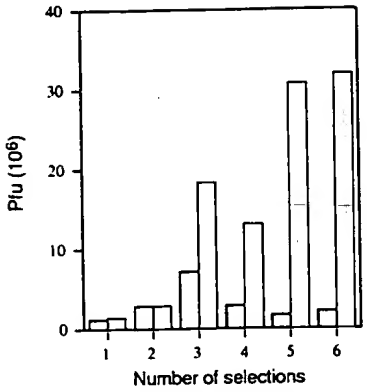


FIG. 4. **Titer of released substrate phages after digestion with rMCP-4.** The C terminus of the capsid protein 10 of the T7 phage was manipulated to contain a nine-amino acid-long random peptide followed by a histidine tag. The histidine tag was used to anchor the phages to a matrix. Phages containing a peptide that is susceptible to protease cleavage are detached from the matrix and can subsequently be removed from the matrix and further propagated. The figure shows the number of plaque forming units (*Pfu*) after each round of selection. The *filled bars* represent library incubated with protease, and the *open bars* represent control library without protease addition.

The C terminus of the capsid protein 10 of the T7 phage was manipulated to contain a nine-amino acid-long random peptide followed by a histidine tag. The histidine tag was used to anchor the phages to a matrix. Phages with a random peptide that is susceptible to protease cleavage will be detached from the matrix and can therefore be removed and be propagated. The release of phages after each protease digestion step was determined by titration and plating in top agarose. The number of cleaved phages after each selection compared with the control is shown in Fig. 4. After six rounds of selection, the addition of rMCP-4 to the amplified library gave a 20-fold higher release of phages than the control library with buffer only.

After six rounds of selection and propagation, 32 individual phage plaques were picked and sequenced (Table I). A consensus could be seen when aligning the phage sequences, especially at the N-terminal side of the scissile bond. The nomenclature by Schechter and Berger (20) is used to designate the amino acid residues in the substrate cleavage region, where P1–P1′ corresponds to the scissile bond. The obtained sequences (indicated in bold type in Table I) were aligned according to the presumed cleavage site. The result for each given subsite is shown in Fig. 5, where the heights of the *bars* in each *diagram* represent the percentages of that amino acid in the subsite, given by the aligned phage display sequences. The predominant P4 amino acid (Fig. 5*A*) is Leu, whereas Val dominates in the P3 position (Fig. 5*B*). A preference for Trp and other aromatic amino acids was seen in P2 (Fig. 5*C*). Only five of the total (32) number of sequenced clones lacked an aromatic P2 amino acid residue, instead they had Leu (two sequences), Met (two sequences), or Ser (one sequence). The phage display analysis indicates that rMCP-4 has a high preference for P1 Phe or Tyr (Fig. 5*D*) and that a broader spectrum of amino acids are tolerated in the P1′ substrate position (Fig. 5*E*). The dominating amino acid in the P2′ position is Gly (47%), *i.e.* an amino acid with a small side chain. Further, several of the other clones recognized by rMCP-4 also had amino acid with relatively small side chains at the P2′ position (Ala, Ser, Fig. 5*F*), indicating a preference for small side chains at this position.

In a control experiment where the phage display library was subjected to cleavage with thrombin, eight clones were sequenced. Four of these clones had the exact same sequence, Thr-Leu-Met-Val-Pro-Arg-Thr-Gly-Ser. When dissecting this sequence into P4–P1, the resulting specificity is P4-Met, P3-Val, P2-Pro, and P1-Arg. The four remaining sequences (the most probable P1 positions are underlined) were Met-Ala-Ser-Pro-Arg-Met-Leu-Arg, Gly-Ala-Lys-Pro-Arg-Ala-Leu-Arg, Ser-Arg-Ala-Ser-Arg-Leu-Lys-Val, and Arg-Trp-Ser-Pro-Arg-Ser-Ile-Tyr-Gly. Using the positional-scanning synthetic combinatorial library method, the substrate specificity for thrombin has been determined previously to P4-Nle (norleucine, substitute for Met), P3-Xaa and P2-Pro (strictly), with the use of P1-Arg (21). Hence, digestion of the phage display library with thrombin resulted in amplification of sequences that are essentially identical to the known cleavage specificity for thrombin. This demonstrates that our T7 phage display substrate library can be used as a powerful tool for determining substrate specificity.

Testing Phage Display Results in a Synthetic Peptide—To verify that the sequences from the phage display analysis were indeed substrates for rMCP-4 and to confirm the cleavage site, a synthetic peptide based on the aligned sequences was prepared. The peptide, NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH was incubated with rMCP-4, and the cleavage products were separated by HPLC. The major peak (and only visible peak except for the starting peptide and the flow-through of Me_2SO) was determined to be NH_2-Leu-Val-Trp-Phe-COOH by mass spectroscopy, demonstrating that the synthetic peptide was cleaved exclusively at the Phe-Arg bond (Fig. 6). The specific activity toward NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH (substrate concentration, 0.1 mM) was 24 pmol of hydrolyzed peptide $s^{-1}/\mu g$ rMCP-4 at a 1:1 protease to heparin (w/w) ratio. Thus, the specific activity of rMCP-4 was ~40-fold higher toward NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH than S-2586 at 1 mM substrate concentration. To determine the K_m and k_{cat} values of rMCP-4 toward NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH, rMCP-4 was incubated with the peptide at increasing concentrations (0.1–1.0 mM), followed by quantitation of formed cleavage products by HPLC analysis as above. Initial experiments showed that incubation of rMCP-4 with NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH (1 mM) gave a linear increase of formed cleavage product up to at 1 h (not shown), and incubations were routinely interrupted within the linear phase of the enzymatic reaction (20 min). Based on these experiments, a K_m value of 1.6 ± 0.2 mM and a k_{cat} value of 1.5 ± 0.13 s^{-1} for NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH were calculated. In contrast with S-2586 (see above), no difference in activity toward NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH could be seen when heparin was excluded.

A search in the SwissProt data base was performed to retrieve potential natural substrate proteins for rMCP-4. A degenerated consensus sequence, based on the results from the phage display experiments was used to perform a search with the ProteinProspector (University of California at San Francisco): P4 (Leu, Val, Trp, Tyr, and Ala), P3 (Val, Thr, Ile,

TABLE I
Target sequences for rMCP-4

Amino acid sequences of the randomized nonamer regions present in rMCP-4 susceptible phages are shown. Susceptible phages were sequenced after six rounds of selection with rMCP-4. Phage clones (32) had their random sequence determined by sequencing with T7Select primers. The general structure of the amino acid sequences in the phage clones is PGGX_9HHHHHH, where *X* indicates the randomized region. The sequences were aligned hypothetically into a P4–P2′ consensus (shown in bold type) where cleavage is proposed to occur between the P1 and P1′ amino acid residue.

Clone no.				P4	P3	P2	P1	P1′	P2′			
12				Y	I	W	F	L	G	E	M	P
13				W	I	W	Y	N	S	V	L	K
2				W	L	F	F	N	G	S	L	W
4				W	V	F	F	S	Y	S	E	L
19				W	V	F	F	Q	S	A	L	P
7				Y	T	Y	Y	S	L	L	V	V
32			G	F	S	F	F	W	G	R	W	
17				I	I	W	Y	V	A	L	P	E
25			G	L	M	F	F	Q	G	G	W	
1				L	M	W	F	R	G	L	S	I
29				L	M	W	F	R	G	A	F	D
15				L	L	Y	F	R	G	V	E	V
18			S	L	V	W	F	R	S	K	F	
26			W	L	V	W	F	R	G	Q	R	
28	R	W	M	L	V	W	F	R	R			
27				V	V	W	F	R	G	V	F	S
11				L	V	Y	Y	V	S	R	R	M
10			G	L	T	W	F	M	G	T	W	
16				I	E	W	Y	S	G	R	L	W
23			A	V	E	Y	Y	R	F	D	S	
3			M	L	T	W	F	F	S	T	P	
8, 21				L	V	W	Y	N	L	P	L	S
14		E	W	M	V	W	F	T	G	V		
9				V	V	W	Y	S	L	V	P	Y
24				V	V	W	Y	V	A	T	L	S
5		D	E	V	V	F	Y	L	G	F		
31			E	L	V	L	F	M	G	G	W	
20				L	V	L	F	A	S	R	V	V
22				L	V	M	F	S	G	T	Y	W
30	D	M	V	W	L	M	Y	R	L			
6	V	A	G	L	G	S	F	G	H			

and Met), P2 (Trp, Phe, and Tyr), P1 (Phe and Tyr), P1′ (Arg, Ser, and Asn), and P2′ (Gly and Leu). This search identified 95 potential human and rodent substrates. Another search was also performed with the sequence P4 (Leu), P3 (Val), P2 (Trp), and P1 (Phe and Tyr) with no P1′ and P2′ amino acid preferences. This search identified 19 proteins. A large number of the identified proteins were regarded as less likely to be substrates, for example if the cleavage sites were in a transmembrane region or if the protein was intracellular. However, among the more likely candidate substrates we identified plasminogen activator inhibitor-1, protein C, procollagen C proteinase enhancer protein, and FcγRIII (Table II).

DISCUSSION

The large amounts of neutral serine proteases synthesized and stored by mast cells suggest an important role for these proteases in mast cell function. Some of the suggested functions of these proteases include activation of matrix metalloproteinases, stimulation of interleukin-8 release along with recruitment of neutrophils, angiotensin I conversion to angiotensin II (human chymase), angiotensin I degradation (rMCP-1), cleavage of vasoactive intestinal peptide, and inactivation of inflammatory neuropeptides (4, 5, 22). Several chymases from different species have been characterized previously, and they demonstrate wide variations with regard to charge, proteoglycan binding properties, catalytic efficiency, and resistance to inhibitors. This indicates that the various related chymases may have different roles in mast cell function. In rodents, the expression pattern for chymases varies with mast cell subtype and location, further supporting distinct functions of the different chymases.

In this study, rat mast cell protease-4 was produced in a mammalian expression system. rMCP-4 is an active protease, as demonstrated by cleavage of a chromogenic chymotrypsin substrate and by binding of [^{3}H]DFP, a potent serine protease inhibitor. Further, the expression system produced essentially 100% correctly folded enzyme, as indicated by active site titration of rMCP-4 with α_1-antichymotrypsin.

Recombinant rMCP-4 bound to heparin with moderate affinity, and this interaction was essential to achieve hydrolysis of S-2586. This may be in agreement with previous findings for rMCP-1, where heparin was shown to stimulate S-2586 hydrolysis by lowering the K_m for this substrate (19). Interestingly, only the cleavage of positively charged substrates, such as S-2586 (Arg at the P3 position) was stimulated by heparin, and it was suggested that heparin acts by masking positively charged groups on the chymase surface, thereby facilitating binding of positively charged substrates. However, whereas the stimulating effect of heparin on rMCP-1 was relatively modest, the presence of heparin is absolutely necessary for detection of rMCP-4-catalyzed hydrolysis of S-2586. This may reflect the presence of positively charged amino acid residues in the vicinity of the active site of rMCP-4, which cause strong repulsion toward positively charged substrates unless neutralized by binding to polyanions such as heparin. It is important to note that even in the presence of heparin, the activity of rMCP-4 is much lower (~700 times) than the activity of rMCP-1 toward the same substrate. This may reflect the highly stringent substrate specificity of rMCP-4 and that the structure of S-2586 differs significantly (except for the P1 position) from the consensus for recognition by rMCP-4 (see below).

Although chymases hydrolyze peptide bonds preferably after Phe or Tyr residues, the extended substrate specificity, and hence the preferred cleavage sites and target protein, can differ. To characterize the extended substrate specificity for

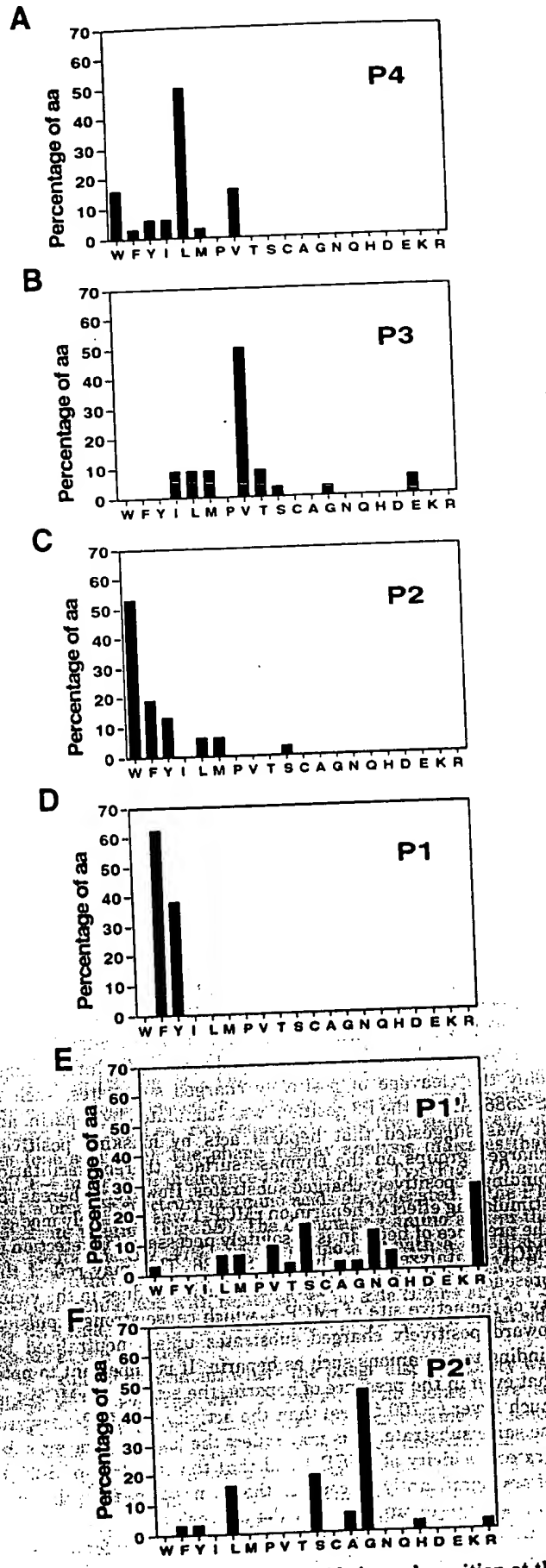


FIG. 5. **The percentage of amino acids in each position at the suggested rMCP-4-susceptible P4-P2' region.** The values were calculated from the aligned phage display sequences shown in Table I.

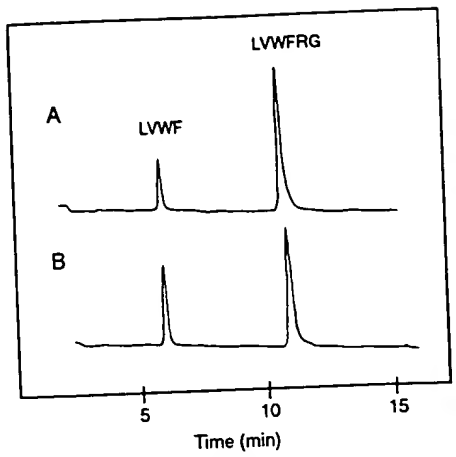


FIG. 6. **Separation of substrate peptide cleavage products by HPLC.** The aligned phage display sequences were used to determine the extended substrate recognition profile for rMCP-4. A synthetic peptide matching this profile was made. The peptide, NH_2-Leu-Val-Trp-Phe-Arg-Gly-COOH (1.04 mM), was incubated with 0.5 μg rMCP-4 and the cleavage products were separated by HPLC. Peak content was determined by mass spectroscopy. One cleavage product, NH_2-Arg-Gly-COOH, was not retained in the column but was eluted with the Me_2SO (the peak corresponding to Me_2SO is not shown). *A*, 1 h of incubation of rMCP-4 with the peptide. *B*, 2 h of incubation with the peptide.

rMCP-4, a library of random peptides displayed on the T phage was screened for protease susceptibility. The phage display analysis revealed that rMCP-4 had surprisingly strict substrate specificity, even when using quite large amounts of protease and relatively long incubation times. The alignment of 32 sequenced phage clones revealed a consensus P4–P2' Leu-Val-Trp-Phe-Arg-Gly. The corresponding peptide was synthesized, and this substrate was hydrolyzed 40 times more efficiently than S-2586 in the presence of heparin.

When comparing the results in the phage display analysis of rMCP-4 with the substrate specificity analysis for rMCP-2 performed by Powers *et al.* (8), several features of these enzymes were shown to be shared. The most preferred P3 amino acid for rMCP-2 was shown to be Val, followed by Leu, Met, Thr, and Phe, a result that correlates well with the phage display result for rMCP-4. Because no P2 aromatic amino acids were included in the rMCP-2 study, it is not possible to determine whether rMCP-2 and rMCP-4 have similar P2 residue preferences. The study by Powers *et al.* (8) also showed that the favored P1 residue was Phe. When substituting P1 Phe with Tyr, a 10-fold decrease of k_{cat}/K_m was observed, and substituting to Trp resulted in a 100-fold lower k_{cat}/K_m. It was recently noted that the bovine chymotrypsin A and B differ in their catalytic activity toward substrates with P1 Trp (23). Bovine chymotrypsin A cleaved substrates with Trp or Phe with the same catalytic activity, whereas the B type chymotrypsin only cleaved P1 Phe substrates efficiently (23). This discrimination of substrates was shown to depend on a single amino acid, Ala^{226} in chymotrypsin B and Gly^{226} in chymotrypsin A. By producing a chymotrypsin B A226G mutant, it was shown that the B type chymotrypsin could gain the same properties as those of the A type, *i.e.* to cleave substrates with P1 Trp. rMCP-4 has a high preference for P1 Phe or Tyr. Chymases generally have an Ala residue at position 226 (chymotrypsin numbering), but an Ile residue occupies position 226 in rMCP-4. This amino acid might restrict the S1 pocket even more and block cleavage against P1 Trp substrates.

In vitro, the substrate specificity of a protease might be well defined. However, *in vivo*, other factors such as localization and the expression level of the protease as well as substrate availability must be taken into consideration. To identify candidate *in vivo* substrates for rMCP-4, data base searches

TABLE II
Potential substrates for rMCP-4

A search in the SwissProt data base was carried out to identify potential substrates for rMCP-4. The search allowed variants in each position, P4 (LVWYA), P3 (VTLM), P2 (WFY), P1 (FY), P1′ (RSN), and P2′ (GLS). The search was limited to human and rodent proteins only and resulted in identification of potential target sequences in 95 proteins. Homologous proteins from all investigated species (human, mouse, and rat) were often identified. However, if a rat protein was found, the human and mouse homologues were not investigated. This limited the number to 47 proteins. Intracellular proteins as well as proteins in which the potential amino acid target sequence was located in a transmembrane domain were deselected as being less likely substrates for rMCP-4. An additional search with P4 (L), P3 (V), P2 (W), and P1 (FY) with no P1′ and P2′ amino acid preferences was also performed. This search identified 19 targets, but after deselection only one potential target protein remained, low affinity Ig γ Fc region receptor III.

Target sequence						
P4	P3	P2	P1	P1′	P2′	Potential substrate, accession number (cleavage site)
V	V	F	F	R	G	Protein C precursor, rat P31394 (406–411)
L	L	W	Y	S	G	Procollagen C-proteinase enhancer protein, rat O08628 (142–147)
V	M	Y	F	N	G	Coagulation factor V, human P12259 (2015–2020)
V	V	Y	Y	N	S	TGF-β receptor type III, rat P26342 (518–523)
V	V	W	Y	S	S	Cysteine-rich secretory protein-3, human P54108 (142–147)
A	L	Y	F	N	G	Plasminogen activator inhibitor-1, rat P20961 (191–196)
L	V	W	F	H	A	Low affinity Ig γ Fc region receptor III, rat P27645 (218–221)

made based on the consensus sequence for substrate recognition by rMCP-4. This strategy identified several potential substrate proteins for rMCP-4. One of these was procollagen C-proteinase enhancer protein. This protein potentiates the activity of procollagen C-proteinase, which removes the C-terminal propeptide of procollagen. The potential cleavage site for rMCP-4 is located in the first (most N-terminal) CUB domain, which is the motif with the enhancing capability. Thus, its cleavage by rMCP-4 may result in decreased amounts of collagen present at the site of mast cell degranulation. Another potential substrate *in vivo* may be plasminogen activator inhibitor-1. Destruction of plasminogen activator inhibitor-1 by rMCP-4 could lead to more plasmin being formed, which in turn may increase the proteolytic degradation of the extracellular matrix and result in increased cell migration and invasion. Further, the serine protease protein C was identified as a potential substrate for rMCP-4. Active protein C functions as a feedback inhibitor in the blood-clotting cascade. Inactivation of protein C could thus result in enhanced coagulation at sites of rMCP-4 release.

The biological function of the MMC proteases is uncertain. However, when we consider the large amounts of proteases stored within the MMC granules and their release under inflammatory conditions, it is likely that they fulfill an important role during MMC-mediated defense reactions in the gut. Direct evidence for involvement of MMC proteases in host defense was recently obtained by using a mouse strain with an engineered deficiency in its expression of mMCP-1 (1). These mice showed a decreased parasite clearance rate after they had been infected with *Trichinella spiralis*, as compared with the normal counterparts. As regards rMCP-4, very little is known about its biological function. It is normally expressed in very low or undetectable levels in the rat intestine. However, when the animals are infected with *Nippostrongylus brasiliensis*, there is a dramatic up-regulation of rMCP-4 mRNA, indicating a role in defense against these parasites. rMCP-4 is most likely expressed by the MMCs, because rMCP-4 was originally cloned from a MMC-like cell line. Moreover, the increase in rMCP-4 expression during *N. brasiliensis* infection correlates well with the up-regulation of the classical MMC protease rMCP-2 that is seen under these circumstances; again consistent with that, rMCP-4 is a true MMC protease. However, it is not known whether the parasite infection causes up-regulated rMCP-4 expression because of increased expression within each cell or whether the increased rMCP-4 levels are due to the increased number of MMCs present.

An identification of the biological role of rMCP-4 or of any other mast cell protease ultimately must involve the determination of exact substrate specificity along with identification of their natural substrates. This is the first report where detailed substrate specificity for a mast cell chymase has been obtained by performing peptide phage display analysis. Future work will be aimed at determining whether potential protein substrates, which have been identified as possible targets based on the presence of consensus sequences recognized by rMCP-4, are also *in vivo* targets for the protease. The peptide phage display technique may also be utilized for determining detailed substrate specificity analysis of other members of the chymase family, which may lead to a better understanding of the role(s) of these enzymes during mast cell-mediated reactions.

Acknowledgment—We are very grateful to Anders Lundequist for performing the HPLC analysis.

REFERENCES

1. Knight, P. A., Wright, S. H., Lawrence, C. E., Paterson, Y. Y., and Miller, H. R. (2000) *J. Exp. Med.* **192**, 1849–1856
2. Mekori, Y. A., and Metcalfe, D. D. (2000) *Immunol. Rev.* **173**, 131–140
3. Miller, H. R. (1996) *Vet. Immunol. Immunopathol.* **54**, 331–336
4. Welle, M. (1997) *J. Leukocyte Biol.* **61**, 233–245
5. Chandrasekharan, U. M., Sanker, S., Glynias, M. J., Karnik, S. S., and Husain, A. (1996) *Science* **271**, 502–505
6. Caughey, G. H., Raymond, W. W., and Wolters, P. J. (2000) *Biochim. Biophys. Acta.* **1480**, 245–257
7. Lutzelschwab, C., Pejler, G., Aveskogh, M., and Hellman, L. (1997) *J. Exp. Med.* **185**, 13–29
8. Powers, J. C., Tanaka, T., Harper, J. W., Minematsu, Y., Barker, L., Lincoln, D., Crumley, K. V., Fraki, J. E., Schechter, N. M., Lazarus, G. G., Nakajima, K., Nakashino, K., Neurath, H., and Woodbury, R. G. (1985) *Biochemistry* **24**, 2048–2058
9. Pemberton, A. D., Huntley, J. F., and Miller, H. R. (1988) *Biochim. Biophys. Acta* **1379**, 29–34
10. Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W., and Neurath, H. (1988) *Biochemistry* **27**, 8097–8105
11. Scudamore, C. L., Jepson, M. A., Hirst, B. H., and Miller, H. R. (1998) *Eur. J. Cell Biol.* **75**, 321–330
12. Lutzelschwab, C., Lunderius, C., Enerback, L., and Hellman, L. (1998) *Eur. J. Immunol.* **28**, 3730–3737
13. Hallgren, J., Karlson, U., Poorafshar, M., Hellman, L., and Pejler, G. (2000) *Biochemistry* **39**, 13068–13077
14. Cwirla, S. E., Peters, E. A., Barrett, R. W., and Dower, W. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6378–6382
15. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7754–7759
16. Trong, H. L., Newlands, G. F., Miller, H. R., Charbonneau, H., Neurath, H., and Woodbury, R. G. (1989) *Biochemistry* **28**, 391–395
17. Sali, A., Matsumoto, R., McNeil, H. P., Karplus, M., and Stevens, R. L. (1993) *J. Biol. Chem.* **268**, 9023–9034
18. Hunt, J. E., Friend, D. S., Gurish, M. F., Feyfant, E., Sali, A., Huang, C., Ghildyal, N., Stechschulte, S., Austen, K. F., and Stevens, R. L. (1997) *J. Biol. Chem.* **272**, 29158–29166
19. Pejler, G., and Sadler, J. E. (1999) *Biochemistry* **38**, 12187–12195
20. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162
21. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nat. Biotechnol.* **18**, 187–193
22. Huang, C., De Sanctis, G. T., O'Brien, P. J., Mizgerd, J. P., Friend, D. S., Drazen, J. M., Brass, L. F., and Stevens, R. L. (2001) *J. Biol. Chem.* **276**, 26278–26284
23. Hudaky, P., Kaslik, G., Venekei, I., and Graf, L. (1999) *Eur. J. Biochem.* **259**, 528–533